BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-88/22

# A MATRIX APPROACH TO LIKELIHOOD EVALUATION AND SIGNAL EXTRACTION FOR ARIMA COMPONENT TIME SERIES MODELS

by

William Bell
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

and

Steven C. Hillmer
301 Summerfield Hall
University of Kansas
Lawrence, KS 66045



Recommended by: Nash J. Monsour

Report completed: August 29, 1988

Report issued: August 29, 1988

# A MATRIX APPROACH TO LIKELIHOOD EVALUATION AND SIGNAL EXTRACTION FOR ARIMA COMPONENT TIME SERIES MODELS

William Bell

Bureau of the Census

Steven C. Hillmer

University of Kansas

June 1988

## 1. Introduction

Three common approaches to evaluating (Gaussian) likelihoods and doing other computations with time series models might be called the classical approach, the Kalman filter approach, and the matrix approach. The classical approach works directly with difference equation forms of models (particularly for autoregressive - integrated - moving average (ARIMA) models) and such things as covariance generating functions and spectral densities. This approach has been used for likelihood evaluation for ARIMA models by Box and Jenkins (1970), Newbold (1974), Dent (1977), Osborn (1977), Hillmer and Tiao (1979), Ljung and Box (1979), Tunnicliffe-Wilson (1983), and others. Spectral approaches to model estimation have also been used; one such reference is Hannan (1970). The classical approach has been used in the signal extraction problem in the stationary case by Whittle (1963), among others, with extensions to the nonstationary case provided by Hannan (1967), Sobel (1967), Cleveland and Tiao (1976), Pierce (1979), and Bell (1984). The Kalman filter approach involves putting the time series model in state-space form and using the Kalman filter in doing likelihood evaluation, as in Gardner, Harvey, and Phillips (1980), R. H. Jones (1980), Pearlman (1980), Kitagawa (1981), Mélard (1984), and others. Signal extraction may be performed with the Kalman smoother as suggested by Pagan (1975), Kitagawa (1981), Burridge and Wallis (1985), and others. The Kalman filter has no inherent limitations to stationary models; however, it does require specification of initial conditions, for which there is typically no basis with nonstationary models involving differencing. Ansley and Kohn (1985) and Kohn and Ansley (1986, 1987) addressed this problem with what they called a "transformation approach" implemented in a modified Kalman filter. Bell and Hillmer (1987a) show how the transformation approach can be implemented with the ordinary Kalman filter. The matrix approach uses matrix results (such as decompositions) to evaluate the determinant and quadratic form in a Gaussian likelihood; this approach can also be used to solve the signal extraction problem. The matrix approach seems to have been less popular in the literature than the classical or Kalman filter approaches. For likelihood evaluation Ansley (1979) suggested use of the Cholesky decomposition, Phadke and Kedem (1978) considered this and a method using Woodbury's formula for the inverse of a matrix of particular form, Wincek and Reinsel (1984) extended the use of the Cholesky decomposition to problems with missing data, and Brockwell and Davis (1987) suggested use of an "innovations algorithm" that amounts to doing a Cholesky decomposition. Carlin (1987) used the sweep operator in a Bayesian analysis involving likelihood evaluation and signal extraction for fractionally intergrated moving average models.

The classical and Kalman filter approaches each have their advantages and disadvantages. Using the classical approach one can easily take advantage of any special structure of the model (such as the multiplicative seasonality of Box and Jenkins (1970)), making this approach convenient and computationally efficient in certain cases. Unfortunately, there are some problems where the classical approach is difficult or impossible to apply, including problems with missing data, variances changing over time, and estimation for component models (one of the problems considered here). Also, finite sample signal extraction requires modifications to the classical results as suggested in Cleveland and Tiao (1976), Bell(1984), and Hillmer (1985). The Kalman filter approach is more general and handles all these problems. Proponents of this approach often cite it for computational efficiency, but some effort may be required to achieve this efficiency because of the large number of zeros in the state space representation of ARIMA models. Also, as a recursive procedure, the Kalman filter gives little insight into the computations, whereas, in the signal extraction problem for example, the classical approach yields filters whose weights can be examined to see the effect of observations in the time series on the signal extraction estimate at a given time point.

Because of the close connection between the Kalman filter and the Cholesky decomposition (Solo (1986) points out that the Kalman filter computes the inverse of the Cholesky factor, and Kohn and Ansley (1984) exploit the connection in using the Kalman filter on seasonal moving average models), the matrix approach can, in principle, be used on any problem on which the Kalman filter approach can be used. The choice between the two could then depend on the ease with which an efficient implementation can be achieved, something that is likely to be problem-dependent. The matrix approach does have one advantage over both the classical and Kalman filter approaches. Results from the classical approach often appear obscure to statisticians who are not time series specialists, and the Kalman filter approach is obscure even to many time series analysts (though this is becoming less so as it becomes better known). Results from the matrix approach should be more accessible to non-time series specialists, and also are more interpretable than those from the Kalman filter approach.

In section 2 of this paper we present our ARIMA component models and assumptions, and section 3 develops matrix results for Gaussian likelihood evaluation for these models. Section 4 develops matrix results for nonstationary signal extraction using the transformation approach of Ansley and Kohn (1985). The matrix results apply the transformation approach directly, rather than implementing it with their modified Kalman filter, or the ordinary Kalman filter with a particular initialization (Bell and Hillmer 1987). In section 5 we show how to compute the matrix results for signal extraction. The approach to computations in sections 3 and 5 uses the Cholesky decomposition approach of Ansley (1979), which applies an autoregressive transformation to the data, allowing the Cholesky decomposition to be taken of a band covariance matrix. Similar ideas could be used for other time series problems, such as forecasting, though we shall not do so here.

## 2. ARIMA Component Models and Assumptions

The general model we shall consider is as follows:

$$Y_t = S_t + N_t \tag{2.1}$$

where the components $S_t$ and $N_t$ follow the ARIMA models

$$\phi_S(B)\delta_S(B)S_t = \theta_S(B)b_t \tag{2.2}$$

$$\phi_N(B)\delta_N(B)N_t = \theta_N(B)c_t. \tag{2.3}$$

Here $\phi_S(B)$, $\delta_S(B)$, etc. are polynomials in the backshift operator B, and $b_t$ and $c_t$ are independent white noise series with variances $\sigma_b^2 > 0$ and $\sigma_c^2 > 0$. For simplicity, we shall assume means are all zero except where stated otherwise. If this is not the case we can simply subtract the means. This general model has wide applicability beyond the classical problem of observations $Y_t$ of a signal $S_t$ that are corrupted by noise (or measurement error) $N_t$. Other applications include seasonal modeling and adjustment ($S_t$ = seasonal, $N_t$ = nonseasonal), model based trend estimation ($S_t$ = trend, $N_t$ = irregular), and periodic sample survey estimation ($S_t$ = true population series, $N_t$ = sampling error).

We shall assume that $\phi_S$ and $\phi_N$ have all zeros outside the unit circle, and $\theta_S$ and $\theta_N$ have all zeros on or outside the unit circle. While $\delta_S$ and $\delta_N$ will most commonly be differencing operators, we do not need to restrict their zeros to the unit circle, and thus can allow for explosive models, or for models with roots outside the unit circle where we do not wish to assume the stationary distribution for the starting values of $S_t$ or $N_t$. We shall assume no common zeros for the pairs $(\phi_S, \theta_S)$, $(\phi_N, \theta_N)$, and $(\phi_S, \phi_N)$, though the last restriction is easily dispensed with. We shall also assume, except where stated otherwise, that $\delta_S$ and $\delta_N$ have no common zeros. This assumption is more key, and different results for signal extraction are developed for a particular case where this does not hold.

Given the above model and assumptions it is well known the observed series $Y_t$ follows the model

$$\phi(B)\delta(B)Y_t = \theta(B)a_t$$

where

$$\phi(B) = \phi_S(B)\phi_N(B), \qquad \delta(B) = \delta_S(B)\delta_N(B),$$

$a_t$ is white noise with variance $\sigma^2 > 0$, and $\theta(B)$ and $\sigma^2$ can be determined from the covariance generating function relation

$$\begin{aligned}\theta(B)\theta(F)\sigma^2 = {} & \phi_N(B)\phi_N(F)\delta_N(B)\delta_N(F)\theta_S(B)\theta_S(F)\sigma_b^2 \\ & + \phi_S(B)\phi_S(F)\delta_S(B)\delta_S(F)\theta_N(B)\theta_N(F)\sigma_c^2\end{aligned} \tag{2.4}$$

where $F = B^{-1}$. The orders of $\phi(B)$, $\delta(B)$, and $\theta(B)$ will be denoted p, d, q, those of $\phi_S$, $\delta_S$, $\theta_S$ denoted $p_S$, $d_S$, $q_S$, and those of $\phi_N$, $\delta_N$, $\theta_N$ denoted $p_N$, $d_N$, $q_N$. (Of course, it is possible for a p, d, or q, to be 0, in which case the corresponding operator is not present in the model, or may be taken as the identity.) We see $p = p_S + p_N$ and $d = d_S + d_N$. It will be convenient to write

$$\delta(B)Y_t = w_t \qquad \delta_S(B)S_t = u_t \qquad \delta_N(B)N_t = v_t.$$

We see that

$$w_t = \delta_N(B)u_t + \delta_S(B)v_t. \tag{2.5}$$

We assume that the series $u_t$, $v_t$, and hence $w_t$ are stationary. This encompasses the assumption on the roots of $\phi_S$ and $\phi_N$, and also an assumption that the starting values for $u_t$ and $v_t$ come from their stationary distribution. We assume that $Y_t$ is observed at time points labelled $t = 1,\ldots,n$. Hence, $w_t$ is available for time points $t = d+1,\ldots,n$. We are thus assuming that there are no missing data. Problems with missing data are typically handled with the Kalman filter, though Wincek and Reinsel (1984) developed a matrix approach that deals with missing data.

The results that follow do not explicitly take account of any multiplicative seasonal structure that may exist in the models for $Y_t$, $S_t$, or $N_t$. It should be obvious how to take advantage of such structure in some of the computations that follow, such as in computing autocovariances. In other computations (see section 3) knowledge of such structure may be of no help. While we are assuming $Var(b_t)$ and $Var(c_t)$ do not depend on t, it is easy to modify our results for the case where they do depend on t, as long as how they do so is known. Finally, we are explicity considering only the case where $Y_t$ is the sum of two component series, but the results extend easily to three or more components.

## 3. Gaussian Likelihood Evaluation

Time series model parameters are frequently estimated by maximizing the Gaussian likelihood function. Here we show how the Gaussian likelihood for the ARIMA component models (2.1) - (2.3) can be evaluated by making an easy extension to the approach using the Cholesky decomposition suggested by Ansley (1979).

The first step is to apply $\delta(B)$ to $Y_t$ to get $w_t$ for $t = d+1,\ldots,n$. Often $\delta(B)$ will be a differencing operator, but it may also include autoregressive parameters to be estimated. This occurs when the model for $S_t$ or $N_t$ has autoregressive term(s) in regard to which we do not wish to assume the stationary distribution for the starting values. We shall use the density of $\underset{\sim}{w} = \underset{\sim}{w}_n^{d+1} = (w_{d+1},\ldots,w_n)^T$ as our likelihood function, a standard procedure that has been justified by Ansley and Kohn (1985). (The superscript T indicates the transpose of a vector or matrix.)

Given $\underset{\sim}{w}$ we make the following transformation suggested by Ansley (1979):

$$z_t = \begin{cases} w_t & t = d+1,\ldots,d+p \\ \phi(B)w_t & t = d+p+1,\ldots,n \end{cases} \tag{3.1}$$

This may be written $\Phi \underset{\sim}{w} = \underset{\sim}{z}$ where

$$\underset{(n-d)\times(n-d)}{\Phi} = \begin{bmatrix} I_p & & & \\ -\phi_p \cdots -\phi_1 & 1 & & \\ \ddots & & \ddots & \\ & -\phi_p \cdots -\phi_1 & & 1 \end{bmatrix} \qquad \underset{\sim}{z} = \begin{bmatrix} z_{d+1} \\ \vdots \\ z_n \end{bmatrix} \tag{3.2}$$

We will need to compute the covariance matrix of $\underset{\sim}{z}$ to get the likelihood.

Since $z_t = w_t$ for $t = d+1,\ldots,d+p$ we need to compute some autocovariances of $w_t = \delta_N(B)u_t + \delta_S(B)v_t$. These can be obtained from those of $u_t$ and $v_t$, which we will also need to be able to compute later for doing signal extraction. Note that $u_t$ and $v_t$ are independent and follow the ARMA models

$$\phi_S(B)u_t = \theta_S(B)b_t \qquad \phi_N(B)v_t = \theta_N(B)c_t.$$

McLeod (1975,1977) gives a method for computing ARMA covariances. To illustrate his approach here let $\gamma_u(k) = \mathrm{Cov}(u_{t-k},u_t)$, and let $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \cdots = \theta_S(B)/\phi_S(B)$ so the $\psi_i$ are obtained by equating coefficients of powers of B in $\phi_S(B)\psi(B) = \theta_S(B)$. Then (let $\theta_{S0} = -1$)

$$\gamma_u(k) - \phi_{S1}\gamma_u(k-1) - \cdots - \phi_{S,p_S}\gamma_u(k-p_S) = (-\theta_{Sk} - \theta_{S,k+1}\psi_1 - \cdots - \theta_{S,q_S}\psi_{q_S-k})\,\sigma_b^2 \tag{3.3}$$

where the right hand side becomes zero for $k > q_S$. Using $\gamma_u(k) = \gamma_u(-k)$ and taking the above equations for $k = 0, 1, \ldots, p_S$ yields $p_S+1$ linear equations which may be solved for $\gamma_u(0), \ldots, \gamma_u(p_S)$. Then $\gamma_u(k)$ for $k = p_S+1, \ldots, p+d_N-1$ may be obtained recursively from (3.3). Similarly, $\gamma_v(k)$ for $k = 0, 1, \ldots, p+d_S-1$ may be obtained. These determine $\gamma_w(k)$ for $k = 0, 1, \ldots, p-1$ through (let $\delta_{N0} = \delta_{S0} = -1$)

$$\gamma_w(k) = \sum_{i=0}^{d_N}\sum_{j=0}^{d_N} \delta_{Ni}\delta_{Nj}\gamma_u(k+i-j) + \sum_{i=0}^{d_S}\sum_{j=0}^{d_S} \delta_{Si}\delta_{Sj}\gamma_v(k+i-j)$$

which determines $\mathrm{Var}(\underset{\sim}{z}_{d+p}^{d+1}) = \mathrm{Var}(\underset{\sim}{w}_{d+p}^{d+1})$.

We also need $\mathrm{Cov}(z_t, z_{t+k})$ for $d+1 \leq t \leq d+p$ and $d+p+1 \leq t+k \leq n$. These are 0 for k > q. Otherwise, we note

$$\begin{aligned}\mathrm{Cov}(z_t,\ z_{t+k}) &= \mathrm{Cov}(w_t,\ \phi(B)w_{t+k}) \\ &= \mathrm{Cov}(\delta_N(B)u_t+\delta_s(B)v_t,\ \phi_N(B)\delta_N(B)\theta_S(B)b_{t+k} + \phi_S(B)\delta_S(B)\theta_N(B)c_{t+k}) \\ &= \mathrm{Cov}(\delta_N(B)u_t,\ \phi_N(B)\delta_N(B)\theta_S(B)b_{t+k}) + \mathrm{Cov}(\delta_S(B)v_t,\ \phi_S(B)\delta_S(B)\theta_N(B)c_{t+k})\end{aligned}$$

These depend only on k, not t. Let $\eta(B) = \phi_N(B)\delta_N(B)\theta_S(B) = 1 + \eta_1 B + \ldots + \eta_m B^m$ where $m = p_N+d_N+q_S$. Now

$$\mathrm{Cov}(u_t,\ b_{t-\ell}) = \begin{cases} 0 & \ell < 0 \\ \psi_\ell \sigma_b^2 & \ell \geq 0 \end{cases}$$

Then (letting $\eta_0 = -1$, and $\sum_{j=k+i}^{m}$ be 0 when k+i > m)

$$\mathrm{Cov}(\delta_N(B)u_t,\ \eta(B)b_{t+k}) = \begin{cases} 0 & k > m \\ \sum_{i=0}^{d_N} \sum_{j=k+i}^{m} \delta_{Ni}\eta_j\psi_{j-(k+i)}\sigma_b^2 & 0 \leq k \leq m \end{cases}$$

We similarly obtain the $\mathrm{Cov}(\delta_S(B)v_t,\ \phi_S(B)\delta_S(B)\theta_N(B)c_{t+k})$, and hence the $\mathrm{Cov}(z_t, z_{t+k})$ needed.

Finally, we see that for t = d+p+1, ..., n

$$z_t = \phi(B)w_t = \theta(B)a_t = \phi_N(B)\delta_N(B)\theta_S(B)b_t + \phi_S(B)\delta_S(B)\theta_N(B)c_t.$$

The two terms on the right hand side are independent moving average series of orders $m = p_N+d_N+q_S$ and $p_S+d_S+q_N$, whose autocovariances are easily computed.

For example, those of $\phi_N(B)\delta_N(B)\theta_S(B)b_t = \eta(B)b_t$ are

$$\mathrm{Cov}(\eta(B)b_t,\ \eta(B)b_{t+k}) = \begin{cases} 0 & k > m \\ \sum_{i=k}^{q} \eta_{i-k}\eta_i\sigma_b^2 & 0 \le k \le m \end{cases}$$

For k = 0, 1, ..., q we add these to the lag k autocovariances of $\phi_S(B)\delta_S(B)\theta_N(B)c_t$ to get the $\mathrm{Cov}(z_t, z_{t+k})$ needed. This is effectively using the covariance generating function (2.4), though we merely wish to compute the autocovariances, we need not solve for $\theta(B)$ and $\sigma^2$.

We have thus shown how to compute all the elements of

$$\Sigma_z \equiv \mathrm{Var}(\underset{\sim}{z}) = \mathrm{Var}(\Phi \underset{\sim}{w}) = \Phi\ \Sigma_w\ \Phi^T. \tag{3.4}$$

Notice that $\Sigma_z$ is a band matrix of bandwidth max(p,q+1), that is, $\mathrm{Cov}(z_i, z_j) = 0$ for $i-j \ge \max(p,q+1)$. Since the Jacobian of the transformation, $|\Phi|$, is 1, the likelihood is the joint density of $\underset{\sim}{z}$:

$$p(\underset{\sim}{z}) = (2\pi)^{-(n-d)/2}\ |\Sigma_z|^{-1/2}\ \exp\{-\ \tfrac{1}{2}\ \underset{\sim}{z}^T\ \Sigma_z^{-1}\ \underset{\sim}{z}\}$$

We thus require the determinant, $|\Sigma_z|$, and the quadratic form, $\underset{\sim}{z}^T\ \Sigma_z^{-1}\ \underset{\sim}{z}$. Following Ansley (1979), we use the Cholesky decomposition of $\Sigma_z$:

$$\Sigma_z = L\ L^T \quad L = [\ell_{ij}] \text{ lower triangular} \tag{3.5}$$

Since $\Sigma_z$ is a band matrix of bandwidth max(p,q+1), so is L, which may be efficiently computed by a routine desinged to take advantage of the band structure (see, e.g. Dongarra et al. (1979)).

We then have

$$|\Sigma_z| = \prod_{i=1}^{n-d} \ell_{ii}^2$$

$$\underset{\sim}{z}^T \Sigma_z^{-1} \underset{\sim}{z} = \underset{\sim}{z}^T (L\ L^T)^{-1} \underset{\sim}{z} = (L^{-1}\underset{\sim}{z})^T (L^{-1}\underset{\sim}{z}) = \sum_{i=1}^{n-d} \epsilon_i^2$$

where $\underset{\sim}{\epsilon} = (\epsilon_1, \ldots, \epsilon_{n-d})^T = L^{-1}\underset{\sim}{z}$, and the $\epsilon_i$ are uncorrelated, unit variance innovations that may be solved for recursively from

$$L\ \underset{\sim}{\epsilon} = \underset{\sim}{z}. \tag{3.6}$$

We could alternatively use the square-root-free Cholesky decomposition $\Sigma_z = L\ D\ L^T$, where D is diagonal and L is unit lower triangular (1's on the diagonal), with obvious modifications to the above. However, the form given above is somewhat more convenient for the signal extraction results later.

The preceeding shows how the likelihood may be evaluated. It may then be maximized by standard numerical techniques to estimate the unknown parameters of $\phi_S(B)$, $\delta_S(B)$, $\theta_S(B)$, $\phi_N(B)$, $\delta_N(B)$, $\theta_N(B)$, $\sigma_b^2$, and $\sigma_c^2$. Care must be taken in doing this to assure that the model is identified, that is, that different values of the parameters do not lead to the same $\Sigma_z$. One possible such problem arises if $\theta_S$ and $\theta_N$ are not restricted in the optimization to have zeros outside or on the unit circle. However, this problem is easily dispensed with without performing restricted optimization -- if the procedure converges to a non-invertible solution (a zero of $\theta_S$ or $\theta_N$ inside the unit circle) one converts this to the corresponding invertible solution (see Box and Jenkins 1970). More serious problems arise if the model is not identified due to an infinite set of combinations of the parameters yielding the same $\Sigma_z$.

We shall not pursue this here; see Hotta (1988) for a discussion of identification of ARIMA component models.

Other refinements to the procedure are possible. A scale constant, e.g. $\sigma_b^2$ or $\sigma_c^2$, may be concentrated out of the likelihood as done by Ansley (1979) for the ARMA model. If the model includes a regression mean function, $\underset{\sim}{X}_t^T \underset{\sim}{\beta}$, for $Y_t$, where $\underset{\sim}{X}_t^T = (X_{1t},\ldots,X_{kt})$ is a k×1 vector of regressors observed at time t and $\underset{\sim}{\beta}$ a k×1 vector of regression parameters, then by taking $Y_t - X_t^T \underset{\sim}{\beta}$, the above procedure yields the likelihood for a given $\underset{\sim}{\beta}$. The joint likelihood may be efficiently maximized by an iterative generalized least squares scheme as suggested by Otto, Bell, and Burman (1987). If $p > q+1$ one could save some computations by taking advantage of the fact that the bandwidth of $\Sigma_z$ is p in the upper left but only q+1 for most of the matrix. This could be done using a backward autoregressive transformation (Ansley 1979) on $w_t$ instead of (3.1). A refinement suggested by Ansley (1979) for multiplicative seasonal ARMA models that takes advantage of zeros within the band structure of $\Sigma_z$ by recognizing corresponding zeros in L does not work here, since it depends on the multiplicative nature of the seasonality and this is lost, in general, with component models.

## 4. Signal Extraction Results

Here we obtain matrix expressions for an estimate $\hat{\underset{\sim}{S}}$, of $\underset{\sim}{S} = (S_1,\ldots,S_n)^T$, and for $\text{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$. The estimate is obtained using the transformation approach of Ansley and Kohn (1985), who develop a modified Kalman filter to calculate the estimate on the grounds that direct calculation of the transformation approach estimate would be difficult. (Bell and Hillmer

(1987a) show how the ordinary Kalman filter may be initialized to yield this estimate.) However, the expressions we give here give some insight into the transformation approach estimate, and in the next section we show how the Cholesky decomposition can be used to compute the estimate and its variance.

The transformation approach estimate of $S_t$ is obtained as follows. Let $\underset{\sim}{S}_* = (S_1,\ldots,S_{d_S})^T$. Following Bell (1984) we can write,

$$S_t = (\underset{\sim}{A}_t^S)^T \underset{\sim}{S}_* + \sum_{i=0}^{t-d_S-1} \xi_i^S u_{t-i} \qquad (4.1)$$

where $\underset{\sim}{A}_t^S$ and the $\xi_i^S$ may be computed from

$$A_{it}^S = \begin{cases} 1 & i=t \\ 0 & i\neq t \end{cases} \qquad t=1,\ldots,d_S \ , \ i=1,\ldots,d_S$$

$$\delta_S(B)\underset{\sim}{A}_t^S = \underset{\sim}{0} \qquad t > d_S$$

$$\delta_S(B)(1 + \xi_1^S B + \xi_2^S B^2 + \ldots) = 1 \ \Rightarrow \ \delta_S(B)\ \xi_i^S = 0 \qquad i > d_S$$

The relation (4.1) also holds for $t=1,\ldots,d_S$ if the sum is interpreted as 0. The transformation approach estimate is found by (1) finding a linear combination of $\underset{\sim}{Y}$, $\underset{\sim}{h}^T\underset{\sim}{Y}$ say, such that $S_t - \underset{\sim}{h}^T\underset{\sim}{Y}$ does not depend on the starting values $\underset{\sim}{S}_*$, and (2) projecting $S_t - \underset{\sim}{h}^T\underset{\sim}{Y}$ on the "differenced data", $\underset{\sim}{w}$, and adding this to $\underset{\sim}{h}^T\underset{\sim}{Y}$. The resulting estimate, $\hat{S}_t$, has error $(S_t - \hat{S}_t)$ that does not depend on $\underset{\sim}{S}_*$ or $\underset{\sim}{N}_* = (N_1,\ldots,N_{d_N})^T$, and has minimum mean squared error

(MMSE) among all linear functions of $\mathbf{Y}$ with this property (Kohn and Ansley (1987)). The estimate is globally optimal, having MMSE among all linear estimators, if $\mathbf{Y}_* = (Y_1, \ldots, Y_d)^T$ is independent of $\{u_t\}$ and $\{v_t\}$ (Assumption A - see Bell and Hillmer (1987a)).

We now give expressions for transformation approach estimates, $\hat{\mathbf{S}}$, of $\mathbf{S}$, and for $\mathrm{Var}(\mathbf{S} - \hat{\mathbf{S}})$, for three different cases regarding $\delta_S(B)$ and $\delta_N(B)$.

## 4.1 Case I: $S_t$ nonstationary, $N_t$ stationary ($\delta_N(B) = 1$)

In this case the transformation approach estimate amounts to using $\mathbf{w}$ to estimate $\mathbf{N} = (N_1, \ldots, N_n)^T$ (call this estimate $\hat{\mathbf{N}}$) and then estimating $\mathbf{S}$ with $\hat{\mathbf{S}} = \mathbf{Y} - \hat{\mathbf{N}}$, and using $\mathrm{Var}(\mathbf{S} - \hat{\mathbf{S}}) = \mathrm{Var}(\mathbf{N} - \hat{\mathbf{N}})$ (since $\mathbf{S} - \hat{\mathbf{S}} = \hat{\mathbf{N}} - \mathbf{N}$).

From (2.5) with $\delta_N(B) = 1$ and $v_t = N_t$ we have $w_t = u_t + \delta_S(B)N_t$, so that $\mathrm{Cov}(w_t, N_j) = \mathrm{Cov}(\delta_S(B)N_t, N_j) = \gamma_N(j-t) - \delta_{S1}\gamma_N(j-t+1) - \cdots - \delta_{S,d_S}\gamma_N(j-t+d_S)$. This yields the elements of $\mathrm{Cov}(\mathbf{w}, \mathbf{N})$ from the autocovariances of $N_t$ which here follows the ARMA model, $\phi_N(B)N_t = \theta_N(B)c_t$. We can write $\mathbf{w} = \mathbf{u} + \Delta_S \mathbf{N}$ so

$$\mathrm{Cov}(\mathbf{w}, \mathbf{N}) = \Sigma_N \Delta_S^T \tag{4.2}$$

where
$$\Delta_S = \begin{bmatrix} -d_{S,d_S} & \cdots & -\delta_{S1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ & & -\delta_{S,d_S} & \cdots & -\delta_{S1} & 1 \end{bmatrix}$$

$\Delta_S$ is an $(n-d)\times n$ matrix that effects differencing by $\delta_S(B)$ (= $\delta(B)$ here).

Then from well-known results on (mean zero) linear projections

$$\hat{\underset{\sim}{N}} = \mathrm{Cov}(\underset{\sim}{w}, \underset{\sim}{N})\, \Sigma_w^{-1}\, \underset{\sim}{w} = \Sigma_N\, \Delta_S^T\, \Sigma_w^{-1}\, \underset{\sim}{w} \qquad (4.3)$$

so that

$$\hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}} = \underset{\sim}{Y} - \Sigma_N\, \Delta_S^T\, \Sigma_w^{-1}\, \underset{\sim}{w} \qquad (4.4)$$

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}}) = \Sigma_N - \Sigma_N\, \Delta_S^T\, \Sigma_W^{-1}\, \Delta_S\, \Sigma_N\, . \qquad (4.5)$$

## 4.2 Case II: $S_t$ and $N_t$ nonstationary, $\delta_S(B)$ and $\delta_N(B)$ have no common zeros

Consider the nonsingular transformation

$$\tilde{\Delta}_S\, \underset{\sim}{S} = \begin{bmatrix} \underset{\sim}{S}_* \\ \underset{\sim}{u} \end{bmatrix} \quad \text{where} \quad \underset{n\times n}{\tilde{\Delta}_S} = \begin{bmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ -\delta_{S,d_S} & \cdots & -\delta_{S1} & 1 & & \\ & \ddots & & \ddots & \ddots & \\ & & -\delta_{S,d_S} & \cdots & -\delta_{S1} & 1 \end{bmatrix} \qquad (4.6)$$

We shall estimate $\underset{\sim}{S}_*$ and $\underset{\sim}{u}$ separately and then for $t > d_S$ use

$\hat{S}_t = \delta_{S1}\hat{S}_{t-1} + \cdots + \delta_{S,d_S}\hat{S}_{t-d_S} + \hat{u}_t$. Also, we shall obtain the error

variance matrix in estimating $[\underset{\sim}{S}_*^T\ \underset{\sim}{u}^T]^T$, and then obtain $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ by inverting the above transformation.

From (2.5) we have

$$\underset{\sim}{w} = \Delta_N\, \underset{\sim}{u} + \Delta_S\, \underset{\sim}{v}$$

where $\Delta_S$ is $(n-d)\times(n-d_N)$ but of the same form as in (4.2) (where $d_N=0$), and $\Delta_N$ is an $(n-d)\times(n-d_S)$ matrix defined analogously to $\Delta_S$ but using $\delta_N(B)$. Then $\mathrm{Cov}(\underset{\sim}{w}, \underset{\sim}{u}) = \Sigma_u \Delta_N^T$, and the elements can be computed from $\mathrm{Cov}(w_t, u_j) = \mathrm{Cov}(\delta_N(B)u_t, u_j) = \gamma_u(j-t) - \delta_{N1}\gamma_u(j-t+1) - \ldots - \delta_{N,d_N}\gamma_u(j-t+d_N)$. Then using $\underset{\sim}{w}$ to estimate $\underset{\sim}{u}$ we have

$$\hat{\underset{\sim}{u}} = \Sigma_u \Delta_N^T \Sigma_w^{-1} \underset{\sim}{w} \qquad \mathrm{Var}(\underset{\sim}{u}-\hat{\underset{\sim}{u}}) = \Sigma_u - \Sigma_u \Delta_N^T \Sigma_w^{-1} \Delta_N \Sigma_u . \tag{4.7}$$

Also estimating $\underset{\sim}{v}$ from $\underset{\sim}{w}$ gives

$$\hat{\underset{\sim}{v}} = \Sigma_v \Delta_S^T \Sigma_w^{-1} \underset{\sim}{w} \qquad \mathrm{Var}(\underset{\sim}{v}-\hat{\underset{\sim}{v}}) = \Sigma_v - \Sigma_v \Delta_S^T \Sigma_w^{-1} \Delta_S \Sigma_v . \tag{4.8}$$

To estimate $\underset{\sim}{S}_*$ we need the following relation between the starting values for $Y_t$ and those for $S_t$ and $N_t$ given by Bell (1984):

$$\underset{\sim}{Y}_* = [H_1 \ H_2] \begin{bmatrix} \underset{\sim}{S}_* \\ \underset{\sim}{N}_* \end{bmatrix} + C_1 \underset{\sim}{u}_d^{d_S+1} + C_2 \underset{\sim}{v}_d^{d_N+1} \tag{4.9}$$

where $\underset{\sim}{u}_d^{d_S+1} = (u_{d_S+1},\ldots,u_d)^T$ , $\underset{\sim}{v}_d^{d_N+1} = (v_{d_N+1},\ldots,v_d)^T$, and

$$\underset{d\times d_S}{H_1} = \begin{bmatrix} I_{d_S} \\ (\underset{\sim}{A}^S_{d_S+1})^T \\ \vdots \\ (\underset{\sim}{A}^S_d)^T \end{bmatrix} \qquad \underset{d\times d_N}{H_2} = \begin{bmatrix} I_{d_N} \\ (\underset{\sim}{A}^N_{d_N+1})^T \\ \vdots \\ (\underset{\sim}{A}^N_d)^T \end{bmatrix}$$

$$\underset{d\times(d-d_S)}{C_1} = \begin{bmatrix} & 0_{d_S\times(d-d_S)} & & \\ \xi_0^S & & & \\ \vdots & \cdot & & \\ & & \cdot & \\ \xi_{d-d_S-1}^S & \ldots & & \xi_0^S \end{bmatrix} \qquad \underset{d\times(d-d_N)}{C_2} = \begin{bmatrix} & 0_{d_N\times(d-d_N)} & & \\ \xi_0^N & & & \\ \vdots & \cdot & & \\ & & \cdot & \\ \xi_{d-d_N-1}^N & \ldots & & \xi_0^N \end{bmatrix}$$

(4.9) just amounts to taking expression (4.1) for $S_t$ and a similar expression for $N_t$ for t = 1,...,d, and adding these together to get $Y_t$ for t = 1,...,d. Bell (1984) observes that the d×d matrix $[H_1\ H_2]$ is nonsingular. We can then obtain from (4.9)

$$\begin{aligned}[I_{d_S}\ 0_{d_S\times d_N}][H_1\ H_2]^{-1}\ \underset{\sim}{Y}_* - \underset{\sim}{S}_* &= [I_{d_S}\ 0_{d_S\times d_N}][H_1\ H_2]^{-1}\{C_1\ \underset{\sim}{u}_d^{d_S+1} + C_2\ \underset{\sim}{v}_d^{d_N+1}\} \\ &= -A_1\ \underset{\sim}{u} - A_2\ \underset{\sim}{v}\end{aligned}$$

where

$$A_1 = -[I_{d_S}\ 0_{d_S\times d_N}][H_1\ H_2]^{-1}\ C_1[I_{d_N}\ 0_{d_N\times(n-d)}]$$
$$A_2 = -[I_{d_N}\ 0_{d_N\times d_S}][H_1\ H_2]^{-1}\ C_2[I_{d_S}\ 0_{d_S\times(n-d)}]$$

The transformation approach estimate of $\underset{\sim}{S}_*$ is

$$\hat{\underset{\sim}{S}}_* = [I_{d_S}\ 0_{d_S\times d_N}][H_1\ H_2]^{-1}\ \underset{\sim}{Y}_* + A_1\ \hat{\underset{\sim}{u}} + A_2\ \hat{\underset{\sim}{v}} \tag{4.10}$$

with error

$$\underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_* = A_1(\underset{\sim}{u} - \hat{\underset{\sim}{u}}) + A_2(\underset{\sim}{v} - \hat{\underset{\sim}{v}}) \tag{4.11}$$

which does not depend on $\underset{\sim}{S}_*$. With $\hat{\underset{\sim}{u}}$ given by (4.7) and $\hat{\underset{\sim}{S}}_*$ by (4.10) we could express $\hat{\underset{\sim}{S}}$ as

$$\hat{\underset{\sim}{S}} = \tilde{\Delta}_S^{-1} \begin{bmatrix} \hat{\underset{\sim}{S}}_* \\ \hat{\underset{\sim}{u}} \end{bmatrix} , \tag{4.12}$$

though it is more convenient to recursively compute $\hat{S}_{d_S+1}, \ldots, \hat{S}_n$ from $\delta_S(B)\hat{S}_t = \hat{u}_t$.

We now obtain $\text{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$. First note that

$$\begin{aligned} \text{Cov}(\underset{\sim}{u} - \hat{\underset{\sim}{u}} , \underset{\sim}{v} - \hat{\underset{\sim}{v}}) &= \text{Cov}(\underset{\sim}{u} - \hat{\underset{\sim}{u}} , \underset{\sim}{v}) \\ &= -\text{Cov}(\hat{\underset{\sim}{u}}, \underset{\sim}{v}) \\ &= -\text{Cov}(\Sigma_u \Delta_N^T \Sigma_w^{-1} \underset{\sim}{w}, \underset{\sim}{v}) \\ &= -\Sigma_u \Delta_N^T \Sigma_w^{-1} \Delta_S \Sigma_v \end{aligned} \tag{4.13}$$

using an orthogonality property of linear projections in the first line and the orthogonality of $\underset{\sim}{u}$ and $\underset{\sim}{v}$ in the second. Then from (4.7), (4.8), (4.11), and (4.13) we have

$$\begin{aligned} \text{Cov}(\underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_*, \underset{\sim}{u} - \hat{\underset{\sim}{u}}) &= A_1 \text{Var}(\underset{\sim}{u} - \hat{\underset{\sim}{u}}) + A_2 \text{Cov}(\underset{\sim}{v} - \hat{\underset{\sim}{v}}, \underset{\sim}{u} - \hat{\underset{\sim}{u}}) \\ &= A_1(\Sigma_u - \Sigma_u \Delta_N^T \Sigma_w^{-1} \Delta_N \Sigma_u) - A_2 \Sigma_v \Delta_S^T \Sigma_w^{-1} \Delta_N \Sigma_u \end{aligned} \tag{4.14}$$

$$\mathrm{Var}(\underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_*) = A_1 \, (\Sigma_u - \Sigma_u \, \Delta_N^T \, \Sigma_w^{-1} \, \Delta_N \, \Sigma_u) \, A_1^T - A_1 \, \Sigma_u \, \Delta_N^T \, \Sigma_w^{-1} \, \Delta_S \, \Sigma_v \, A_2^T$$
$$- A_2 \, \Sigma_v \, \Delta_S^T \, \Sigma_w^{-1} \, \Delta_N \, \Sigma_u \, A_1^T + A_2 \, (\Sigma_v - \Sigma_v \, \Delta_S^T \, \Sigma_w^{-1} \, \Delta_S \, \Sigma_v) \, A_2^T \, . \qquad (4.15)$$

(4.7), (4.14), and (4.15) complete the specification of the error variance matrix of $[\hat{\underset{\sim}{S}}_*^T \;\; \hat{\underset{\sim}{u}}^T]^T$; then from (4.6) and (4.12) we obtain

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \tilde{\Delta}_S^{-1} \, \mathrm{Var}\left[\begin{bmatrix} \underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_* \\ \underset{\sim}{u} - \hat{\underset{\sim}{u}} \end{bmatrix}\right] \tilde{\Delta}_S^{-T} \qquad (4.16)$$

where $\tilde{\Delta}_S^{-T}$ denotes the inverse of $\Delta_S^T$ .

### 4.3 Case III: $S_t$ and $N_t$ nonstationary, $\delta_S(B)$ and $\delta_N(B)$ have common zero(s), $\mathrm{Var}(\underset{\sim}{N}_*)$ known, and $(\underset{\sim}{N}_*)$ independent of $\{u_t\}$ and $\{v_t\}$

Component models where $\delta_S(B)$ and $\delta_N(B)$ have a common zero have been used in a seasonal adjustment context by Cleveland and Tiao (1976) and Burridge and Wallis (1985), but seasonal modeling or adjustment is not the application we have in mind here. In fact, arguments can be made against $\delta_S(B)$ and $\delta_N(B)$ having a common zero in this context (see Bell and Hillmer 1984), and also estimation of $S_t$ and $N_t$ when $\delta_S(B)$ and $\delta_N(B)$ have a common zero requires assumptions about starting values such as those in this subsection's heading or others (see, e.g., Kohn and Ansley (1987)), for which there is generally little basis in seasonal modeling.

The application we have in mind here is estimation in periodic surveys where $S_t$ represents the true underlying series and $N_t$ the sampling error (see

Scott and Smith (1974), R. G. Jones (1980), and Bell and Hillmer (1987b)). Typically $S_t$ will require differencing, and one can conceive of situations where the model for $N_t$ might also involve differencing. This could arise if a nonstationary model was used to explain correlation over time for units in the population being sampled, and the sample design were such that the resulting $N_t$ followed, at least approximately, a nonstationary model. This might happen in a panel study where units remain in sample a long time, or even indefinitely. Since we should have available an estimate of $\text{Var}(\underset{\sim}{N}_*)$ in this case, we have the situation we shall consider here if $\underset{\sim}{N}_*$ is independent of $\{u_t\}$ and $\{v_t\}$. This last assumption may be more open to question, but could be considered with regard to any particular application, or perhaps the results given here can be modified. We should point out that we have not actually attempted nonstationary modeling of sampling error - modeling of time series subject to sampling error being still in its infancy - but are presenting results here that may be used in this case should such a model be developed.

If $\delta_S(B)$ and $\delta_N(B)$ have common zero(s) we write

$$\delta(B) = \delta_S^*(B)\delta_N^*(B)\delta_c(B) \qquad \delta_c(B) = 1 - \delta_{c1}B - \dots - \delta_{c,d_c}B^{d_c}$$

where $\delta_c(B)$ is the product of the $d_c$ common factors in $\delta_S(B)$ and $\delta_N(B)$, $\delta_S^*(B) = \delta_S(B)/\delta_c(B)$, $\delta_N^*(B) = \delta_N(B)/\delta_c(B)$, and $d = d_S + d_N - d_c$. (Actually, the approach taken here seems most appropriate when $\delta_S(B)$ contains $\delta_N(B)$, so $\delta_c(B) = \delta_N(B)$, $\delta_N^*(B) = 1$, and $d = d_S$. If this is not the case, part of the effect of the starting values $\underset{\sim}{N}_*$ can be eliminated, which may yield better results than those presented here.) We then have

$$w_t = \delta(B)Y_t = \delta_N^*(B)u_t + \delta_S^*(B)v_t$$

$$\underset{\sim}{w} = \Delta_N^* \underset{\sim}{u} + \Delta_S^* \underset{\sim}{v}$$

where $\Delta_N^*$ is an $(n-d)\times(n-d_S)$ matrix corresponding to $\delta_N^*(B)$, and $\Delta_S^*$ is an $(n-d)\times(n-d_N)$ matrix corresponding to $\delta_S^*(B)$, analogous to (4.2). Notice that

$$\underset{\sim}{Y}^1_{d_S} = \underset{\sim}{S}_* + \underset{\sim}{N}^1_{d_S}$$

The transformation approach will eliminate the effects of $\underset{\sim}{S}_*$, but not of $\underset{\sim}{N}_*$, since when $\delta_S(B)$ and $\delta_N(B)$ have common zeroes we cannot eliminate both $\underset{\sim}{S}_*$ and $\underset{\sim}{N}_*$.

First, consider the case $d_N \geq d_S$ so that $\underset{\sim}{N}^1_{d_S}$ is part of $\underset{\sim}{N}_* = \underset{\sim}{N}^1_{d_N}$. Note

$$\text{Cov}(\underset{\sim}{N}_*, \underset{\sim}{w}) = \text{Cov}(\underset{\sim}{N}^*, \Delta_N^* \underset{\sim}{u} + \Delta_S^* \underset{\sim}{v}) = 0$$

so the estimate of $\underset{\sim}{N}_*$ using $\underset{\sim}{w}$ is $\underset{\sim}{0}$. The transformation approach then uses

$$\hat{\underset{\sim}{S}}_* = \underset{\sim}{Y}^1_{d_S} \qquad \hat{\underset{\sim}{u}} = \Sigma_u \Delta_N^* \Sigma_W^{-1} \underset{\sim}{w}$$

and $\delta_S(B)\hat{S}_t = \hat{u}_t$ for $t = d_S+1, \ldots, n$. We also have

$$\mathrm{Var}(\underset{\sim}{u} - \hat{\underset{\sim}{u}}) = \Sigma_u - \Sigma_u (\Delta_N^*)^T \Sigma_w^{-1} \Delta_N^* \Sigma_u$$

$$\mathrm{Var}(\underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_*) = \mathrm{Var}(-\underset{\sim}{N}^1_{d_S}) = d_S x d_S \text{ upper left-hand corner of } \mathrm{Var}(\underset{\sim}{N}_*)$$

$$\mathrm{Cov}(\underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_*, \underset{\sim}{u} - \hat{\underset{\sim}{u}}) = \mathrm{Cov}(-\underset{\sim}{N}^1_{d_S}, \underset{\sim}{u} - \hat{\underset{\sim}{u}}) = 0 \ .$$

With these pieces we can obtain $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ using (4.16).

If $d_S > d_N$ we estimate $\underset{\sim}{N}_*$ by $\hat{\underset{\sim}{N}}_* = \underset{\sim}{0}$ again, and then estimate $\underset{\sim}{v}$ by

$$\hat{\underset{\sim}{v}} = \Sigma_v \Delta_S^* \Sigma_w^{-1} \underset{\sim}{w} \ .$$

After recursively computing $\hat{N}_t = \delta_{N1}\hat{N}_{t-1} + \cdots + \delta_{N,d_N}\hat{N}_{t-d_N} + \hat{v}_t$ for $t = d_N+1,\ldots,n$, we compute $\hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}}$. Also $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}})$, and the latter may be obtained from

$$\mathrm{Var}(\underset{\sim}{N}_* - \hat{\underset{\sim}{N}}_*) = \mathrm{Var}(\underset{\sim}{N}_*) \qquad \text{(assumed known)}$$

$$\mathrm{Cov}(\underset{\sim}{N}_* - \hat{\underset{\sim}{N}}_*, \underset{\sim}{v} - \hat{\underset{\sim}{v}}) = \mathrm{Cov}(\underset{\sim}{N}_*, \underset{\sim}{v} - \hat{\underset{\sim}{v}}) = 0$$

$$\mathrm{Var}(\underset{\sim}{v} - \hat{\underset{\sim}{v}}) = \Sigma_v - \Sigma_v (\Delta_S^*)^T \Sigma_w^{-1} \Delta_S^* \Sigma_v$$

and

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}}) = \tilde{\Delta}_N^{-1} \mathrm{Var}\left( \begin{bmatrix} \underset{\sim}{N}_* \\ \underset{\sim}{v} \end{bmatrix} - \begin{bmatrix} \hat{\underset{\sim}{N}}_* \\ \hat{\underset{\sim}{v}} \end{bmatrix} \right) \tilde{\Delta}^{-T}$$

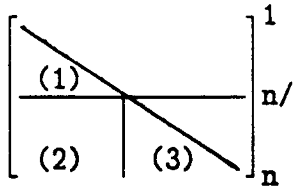
where

$$\tilde{\Delta}_N = \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ -\delta_{N,d_N} & \cdots & -\delta_{N1} & 1 & & & \\ & \ddots & & \ddots & \ddots & & \\ & & -\delta_{N,d_N} & \cdots & -\delta_{N1} & 1 \end{bmatrix} .$$

5. Signal Extraction Computations

We now show how to efficiently compute $\hat{\underset{\sim}{S}}$ and $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ given by the expressions in section 4. Along with specific schemes for each of the three cases, there are some general considerations for computational efficiency. One is that the roles of $S_t$ and $N_t$ are interchangeable, that is, instead of directly computing $\hat{\underset{\sim}{S}}$ and $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$, we can compute $\hat{\underset{\sim}{N}}$ and $\mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}})$, and then use $\hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}}$ and $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}})$ (since $\underset{\sim}{S} - \hat{\underset{\sim}{S}} = \underset{\sim}{N} - \hat{\underset{\sim}{N}}$). This fact was already used in sections 4.1 and 4.3. For the case considered in section 4.2 it will generally be easier to compute $\hat{\underset{\sim}{S}}$ and $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ as given there if $d_S < d_N$, and easier to compute the corresponding results for $\hat{\underset{\sim}{N}}$ and $\mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}})$ if $d_S > d_N$. Here we shall show how to compute the results specifically given in section 4.

Other general computational savings are possible in computing $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$. First, and most obvious, since $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ is symmetric, it is determined by its lower triangle. Second, (ignoring the case in section 4.3 for the moment), since the models (2.1) - (2.3) hold for the series reversed in time, i.e. t running from n to 1, it follows that $\mathrm{Var}(S_t - \hat{S}_t) = \mathrm{Var}(S_{n+1-t} - \hat{S}_{n+1-t})$, with analogous results for covariances. If we partition $\mathrm{Var}(\underset{\sim}{S} - \underset{\sim}{\hat{S}})$ as

$\mathrm{Var}(\underset{\sim}{S} - \underset{\sim}{\hat{S}}) =$ 

(or (n+1)/2 for n odd)

then (3) is the transpose of the mirror image of (1) , so the elements in (1) and (2) are sufficient to determine $\mathrm{Var}(\underset{\sim}{S} - \underset{\sim}{\hat{S}})$. This does not hold for the case covered in section 4.3 because of the special assumptions about $\underset{\sim}{N}_*$, which appears in $\underset{\sim}{S} - \underset{\sim}{\hat{S}}$ . If $\Omega = [\omega_{ij}] = \mathrm{Var}(\underset{\sim}{S} - \underset{\sim}{\hat{S}})$ these two restrictions mean that we only need $\omega_{ij}$ for $i \geq j$ and $j \leq (n+1)/2$ to determine $\Omega$. Finally, it will be rare that all of $\mathrm{Var}(\underset{\sim}{S} - \underset{\sim}{\hat{S}})$ will be of interest, at least for n reasonably large. For example, $\mathrm{Cov}(S_1 - \hat{S}_1, S_n - \hat{S}_n)$ will rarely be needed. This makes possible some significant computational savings for the case of section 4.1; these will be outlined in section 5.1.

The basis for our computation schemes here is the computation and Cholesky decomposition of $\Sigma_z$ discussed in section 3. Thus, we start from (see (3.4) and (3.5))

$$\Sigma_z = \Phi \, \Sigma_w \, \Phi^T = L \, L^T$$

where $\Sigma_z$ and the Cholesky factor L are band matrices. From this we have

$$\Sigma_w^{-1} = \Phi^T \Sigma_z^{-1} \Phi = \Phi^T L^{-T} L^{-1} \Phi \tag{5.1}$$

All our estimates here involve

$$\Sigma_w^{-1} \underset{\sim}{w} = \Phi^T L^{-T} L^{-1} \underset{\sim}{z} = \Phi^T L^{-T} \underset{\sim}{\epsilon}$$

where $\underset{\sim}{\epsilon} = \underset{\sim}{L}^{-1} \underset{\sim}{z}$ may be solved for recursively from $L \underset{\sim}{\epsilon} = \underset{\sim}{z}$. Letting $\underset{\sim}{r} = L^{-T} \underset{\sim}{\epsilon}$ we may solve recursively for its elements $r_{n-d}, \ldots, r_1$ (bottom to top) from $L^T \underset{\sim}{r} = \underset{\sim}{\epsilon}$. We can then easily compute $\Sigma_w^{-1} \underset{\sim}{w} = \Phi^T \underset{\sim}{r}$, ignoring the zeros in $\Phi^T$ when taking this product, though some of the approaches that follow do not explicitly compute this last product.

## 5.1 Computing Results for Case I ($N_t$ Stationary)

Recall that we wish to compute, from (4.3) - (4.5)

$$\hat{\underset{\sim}{N}} = \Sigma_N \Delta_S^T \Sigma_w^{-1} \underset{\sim}{w} \qquad \hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}}$$

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \mathrm{Var}(\underset{\sim}{N} - \hat{\underset{\sim}{N}}) = \Sigma_N - \Sigma_N \Delta_S^T \Sigma_w^{-1} \Delta_S \Sigma_N$$

We discussed computation of ARMA covariances in section 3. For the results here we need $\gamma_N(0), \ldots, \gamma_N(n-1)$ to determine $\Sigma_N$. Then $\hat{\underset{\sim}{N}}$ can be computed by computing $\Sigma_w^{-1} \underset{\sim}{w}$ as described above, multiplying this by $\Delta_S^T$ taking account of the many zeroes in $\Delta_S^T$, and then multiplying this result by $\Sigma_N$. That is

$$\hat{\underset{\sim}{N}} = \Sigma_N (\Delta_S^T (\Sigma_w^{-1} \underset{\sim}{w})) = \Sigma_N (\Delta_S^T (\Phi^T (L^{-T} \underset{\sim}{\epsilon}))) \tag{5.2}$$

where the parentheses indicate the order of computation.

If we are also going to compute $\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$, we can use an alternative aproach to computing $\hat{\underset{\sim}{N}}$. From (5.1) and (5.2) we can write

$$\hat{\underset{\sim}{N}} = (\Phi\ \Delta_S\ \Sigma_N)^T\ (L^{-T}\ \underset{\sim}{\epsilon}) = [L^{-1}\ (\Phi\ \Delta_S\ \Sigma_N)]^T\ \underset{\sim}{\epsilon} \qquad (5.3)$$

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \Sigma_N - [L^{-1}\ (\Phi\ \Delta_S\ \Sigma_N)]^T\ [L^{-1}\ (\Phi\ \Delta_S\ \Sigma_N)] \qquad (5.4)$$

Obviously, we need to compute $\Phi\ \Delta_S\ \Sigma_N$ and then $L^{-1}\ (\Phi\ \Delta_S\ \Sigma_N)$. We start with just $\Delta_S\ \Sigma_N$. The $ij^{th}$ element of $\Delta_S\ \Sigma_N$ is $a_{i-j}$ where the sequence $a_k$ is defined by

$$a_k \equiv \mathrm{Cov}(\delta_S(B)N_t,\ N_{t+k}) = \gamma_N(k) - \delta_{S1}\gamma_N(k+1) - \dots - \delta_{S,d_S}\gamma_N(k+d_S).$$

Then $a_{d_N+1-n},\dots,a_{n-d-1}$ determine $\Delta_S\ \Sigma_N$ through

$$\underset{(n-d)\times(n-d_N)}{\Delta_S\ \Sigma_N} = \begin{bmatrix} a_{-d_S} & a_{1-d_S} & \cdots & a_{n-d-1} \\ a_{-d_S-1} & a_{-d_S} & \ddots & \\ \vdots & \ddots & \ddots & \\ a_{d_N+1-n} & & & \end{bmatrix}$$

(We are showing the indices here in a way that the results can be easily used in the next section where $v_t$ replaces $N_t$, or $u_t$ and $\Delta_N$ replace $N_t$ and $\Delta_S$. Recall that here $d_N = 0$ and $d_S = d$.) Notice $\Delta_S\ \Sigma_N$ is determined by its first row and column. If there are no AR operators, i.e. $\phi_S(B) = \phi_N(B) = 1$, we replace $\Phi$ by the identity matrix. Otherwise, consider $\Phi$ given in (3.2). We see the first p rows of $\Phi\ \Delta_S\ \Sigma_N$ are just those of $\Delta_S\ \Sigma_N$. The remaining elements are determined by the sequence $\lambda_k$ defined by

$$\lambda_k \equiv \mathrm{Cov}(\phi(B)\delta_S(B)N_t,\ N_{t+k}) = a_k - \phi_1 a_{k+1} - \cdots - \phi_p a_{k+p} \ .$$

We compute $\lambda_{d_N+1-n},\ldots,\lambda_{n-d-p-1}$ and then

$$\Phi\ \Delta_S\ \Sigma_N = \left[\begin{array}{cccc} a_{-d_S} & a_{1-d_S} & \cdots & a_{n-d-1} \\ \vdots & \ddots & & \ddots \\ a_{1-d_S-p} & & & \\ \hline \lambda_{-d_S-p} & \lambda_{1-d_S-p} & \cdots & \lambda_{n-d-p-1} \\ \lambda_{-d_S-p-1} & \lambda_{-d_S-p} & \ddots & \\ \vdots & \ddots & & \ddots \\ \lambda_{d_N+1-n} & & & \end{array}\right] \begin{array}{l} \left\{ \begin{array}{l}\text{first } p \text{ rows} \\ \text{of } \Delta_S\ \Sigma_N\end{array}\right. \\ \\ \\ \\ \\ \\ \end{array} \tag{5.5}$$

Notice this matrix is determined by its first and $p^{th}$ rows, and first column.

Having thus computed $\Phi\ \Delta_S\ \Sigma_N$ we can then compute

$$R = L^{-1}\ (\Phi\ \Delta_S\ \Sigma_N) \quad \Rightarrow \quad L\ R = \Phi\ \Delta_S\ \Sigma_N \ . \tag{5.6}$$

We solve the second relation recursively for each column $\underset{\sim}{r}_i$ of $R = [\underset{\sim}{r}_1,\ldots,\underset{\sim}{r}_n]$. We then have from (5.2), (5.3), and (5.6)

$$\hat{\underset{\sim}{N}} = R^T\ \underset{\sim}{\epsilon} \qquad \hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}}$$

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \Sigma_N - R^T\ R\ . \tag{5.7}$$

Actually, since it is easy to solve $L^T\underset{\sim}{r} = \underset{\sim}{\epsilon}$ for $\underset{\sim}{r} = L^{-T}\underset{\sim}{\epsilon}$ , we may prefer to compute $\hat{\underset{\sim}{N}}$ as $\hat{\underset{\sim}{N}} = (\Phi\ \Delta_S\ \Sigma_N)^T\ \underset{\sim}{r}$ .

We will seldom need all of $\text{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$, and it is easy to use (5.7) to compute only those elements needed. If $\Omega = [\omega_{ij}] = \text{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ then

$$\omega_{ij} = \gamma_N(i - j) - \underset{\sim}{r}_i^T\ \underset{\sim}{r}_j\ .$$

We can save considerable computations by computing only those $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_j$ needed. If we only want $\text{Var}(S_t - \hat{S}_t)$ for $t = 1,\ldots,n$ we only compute the required $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_i$ . If we also want $\text{Cov}(S_t - \hat{S}_t,\ S_{t-1} - \hat{S}_{t-1})$ we also compute $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_{i-1}$ as required, and so on. (See also the discussion at the beginning of this section on what computations are required.) We can also save on computer storage with this approach since as we sequentially compute $\underset{\sim}{r}_i$ for $i = 1, 2, \ldots$, we can compute $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_i$, $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_{i-1}$, $\underset{\sim}{r}_i^T\ \underset{\sim}{r}_{i-2}$, etc. as desired; and then discard the $\underset{\sim}{r}_{i-j}$ as they are not needed (i.e. large j).

### 5.2 Computing Results for Case II ($S_t$, $N_t$ nonstationary, $\delta_S(B)$ and $\delta_N(B)$ have no common zeros)

To produce the estimate $\hat{\underset{\sim}{S}}$ we need to compute (see (4.7), (4.8), (4.10))

$$\hat{\underset{\sim}{u}} = \Sigma_u\ \Delta_N^T\ \Sigma_w^{-1}\ \underset{\sim}{w} = (\Phi\ \Delta_N\ \Sigma_u)^T\ \underset{\sim}{r} \quad \text{where } \underset{\sim}{r} = L^{-T}\ \underset{\sim}{\epsilon}$$

$$\hat{\underset{\sim}{v}} = \Sigma_v\ \Delta_S^T\ \Sigma_w^{-1}\ \underset{\sim}{w} = (\Phi\ \Delta_S\ \Sigma_v)^T\ \underset{\sim}{r}$$

$$\hat{S}_* = [I_{d_S} \; 0_{d_S \times d_N}] \; [H_1 \; H_2]^{-1} \; Y_* + A_1 \; \hat{u} + A_2 \; \hat{v}$$

We compute $\Sigma_u$, $\Sigma_v$, $\Phi \Delta_N \Sigma_u$ , $\Phi \Delta_S \Sigma_v$ , $r$, and hence $\hat{u}$ and $\hat{v}$ as discussed earlier (sections 3 and 5.1). We actually do not need all of $\hat{v}$, as will be seen shortly. To compute $\hat{S}_*$ first notice its first term may be computed directly noting that $[I_{d_S} \; 0_{d_S x d_N}] \; [H_1 \; H_2]^{-1}$ is the first $d_S$ rows of $[H_1 \; H_2]^{-1}$. Next let

$$B_1 = - [I_{d_S} \; 0_{d_S x d_N}] \; [H_1 \; H_2]^{-1} \; C_1 \qquad \text{so} \qquad A_1 = B_1 \; [I_{d_N} \; 0_{d_N x (n-d)}]$$

$$B_2 = - [I_{d_S} \; 0_{d_S x d_N}] \; [H_1 \; H_2]^{-1} \; C_2 \qquad \text{so} \qquad A_2 = B_2 \; [0_{d_S} \; 0_{d_S x (n-d)}]$$

These are the first $d_S$ rows of $-[H_1 \; H_2]^{-1} \; C_1$ and $-[H_1 \; H_2]^{-1} \; C_2$, respectively, and these products can be taken directly, taking account of the fact that the first $d_S$ rows of $C_1$ and the first $d_N$ rows of $C_2$ are 0. Then

$$(\Phi \; \Delta_N \; \Sigma_u) \; A_1^T = (\Phi \; \Delta_N \; \Sigma_u) \begin{bmatrix} I_{d_N} \\ 0_{(n-d) x d_N} \end{bmatrix} B_1^T = \begin{bmatrix} \text{first } d_N \text{ columns} \\ \text{of } \Phi \; \Delta_N \; \Sigma_u \end{bmatrix} B_1^T$$

$$(\Phi \; \Delta_S \; \Sigma_v) \; A_2^T = (\Phi \; \Delta_S \; \Sigma_v) \begin{bmatrix} I_{d_S} \\ 0_{(n-d) x d_S} \end{bmatrix} B_2^T = \begin{bmatrix} \text{first } d_s \text{ columns} \\ \text{of } \Phi \; \Delta_S \; \Sigma_v \end{bmatrix} B_2^T.$$

We can then directly compute

$$A_1 \hat{\underset{\sim}{u}} = [(\Phi\ \Delta_N\ \Sigma_u)\ A_1^T]^T\ \underset{\sim}{r} \qquad A_2 \hat{\underset{\sim}{v}} = [(\Phi\ \Delta_S\ \Sigma_v)\ A_2^T]^T\ \underset{\sim}{r}$$

and then compute $\hat{\underset{\sim}{S}}_*$. Notice that since we need only $A_2 \hat{\underset{\sim}{v}}$ and not all of $\hat{\underset{\sim}{v}}$, we need only the first $d_S$ columns of $\Phi\ \Delta_S\ \Sigma_v$, which are determined by the first column and first $d_S$ elements of the first and $p+1^{st}$ rows (see (5.6)).

Having computed $\hat{\underset{\sim}{S}}_*$ and $\hat{\underset{\sim}{u}}$, we obtain $\hat{S}_t$ for $t = d_S+1, \ldots, n$ from

$$\hat{\underset{\sim}{S}}_t = \delta_{S1}\ \hat{S}_{t-1} + \cdots + \delta_{S,d_S}\ \hat{S}_{t-d_S} + \hat{u}_t \qquad t = d_S+1, \ldots, n$$

To compute $\Omega = \mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$ we first compute $\tilde{\Omega} = \mathrm{Var}\begin{bmatrix} \underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_* \\ \underset{\sim}{u} - \hat{\underset{\sim}{u}} \end{bmatrix}$. From (4.2) we have that

$$\begin{bmatrix} \underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_* \\ \underset{\sim}{u} - \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ I_{n-d_S} & 0 \end{bmatrix} \begin{bmatrix} \underset{\sim}{u} - \hat{\underset{\sim}{u}} \\ \underset{\sim}{v} - \hat{\underset{\sim}{v}} \end{bmatrix} \tag{5.8}$$

(4.7), (4.8), and (4.12) can be re-expressed as

$$\mathrm{Var}\begin{bmatrix} \underset{\sim}{u} - \hat{\underset{\sim}{u}} \\ \underset{\sim}{v} - \hat{\underset{\sim}{v}} \end{bmatrix} = \begin{bmatrix} \Sigma_u & 0 \\ 0 & \Sigma_v \end{bmatrix} - \begin{bmatrix} \Sigma_u\ \Delta_N^T \\ \Sigma_v\ \Delta_S^T \end{bmatrix} \Sigma_W^{-1}\ [\Delta_N\ \Sigma_u \mid \Delta_S\ \Sigma_v] \tag{5.9}$$

Using (5.8) and (5.9) and simplifying we eventually get

$$\tilde{\Omega} = \mathrm{Var}\begin{bmatrix} \underset{\sim}{S}_* - \hat{\underset{\sim}{S}}_* \\ \underset{\sim}{u} - \hat{\underset{\sim}{u}} \end{bmatrix} = \begin{bmatrix} A_1 \Sigma_u A_1^T + A_2 \Sigma_v A_2^T & A_1 \Sigma_u \\ \Sigma_u A_1^T & \Sigma_u \end{bmatrix} - [R_1 \; R_2]^T [R_1 \; R_2] \quad (5.10)$$

where $[R_1 \; R_2]$ is $(n-d)\times n$ with

$$R_1 = L^{-1} (\Phi \, \Delta_N \Sigma_u A_1^T + \Phi \, \Delta_S \Sigma_v A_2^T) \quad \text{is} \quad (n-d)\times d_S$$

$$R_2 = L^{-1} (\Phi \, \Delta_N \Sigma_u) \quad \text{is} \quad (n-d)\times(n-d_S)$$

Given that we have computed $\Sigma_u$, $\Sigma_v$, $\Phi \, \Delta_N \Sigma_u$ , $\Phi \, \Delta_N \Sigma_u A_1^T$, and $\Phi \, \Delta_S \Sigma_v A_2^T$ , we show how to compute the rest of the quantities needed.

1) $$A_1 \Sigma_u = B_1 \, [I_{d_N} \; 0_{d_N x(n-d)}] \, \Sigma_u = B_1 \begin{bmatrix} \gamma_u(0) & \cdots & \gamma_u(d_S+1-n) \\ \vdots & & \vdots \\ \gamma_u(d_N-1) & \cdots & \gamma_u(d-n+1) \end{bmatrix}$$

$$A_1 \Sigma_u A_1^T = (A_1 \Sigma_u) \begin{bmatrix} I_{d_N} \\ 0_{(n-d)\times d_N} \end{bmatrix} B_1^T = \begin{bmatrix} \text{first } d_N \text{ columns} \\ \text{of } A_1 \Sigma_U \end{bmatrix} B_1^T$$

$$A_2 \Sigma_v A_2^T = B_2 [I_{d_S} \; 0_{d_S\times(n-d)}] \Sigma_v \begin{bmatrix} I_{d_S} \\ 0_{(n-d)xd_S} \end{bmatrix} B_2^T = B_2 \begin{bmatrix} \gamma_v(0) & & \\ \vdots & \ddots & \\ \gamma_v(d_N-1) & \cdots & \gamma_v(0) \end{bmatrix} B_2^T$$

Compute all these directly.

2) Compute $R_1$ and $R_2$ by solving recursively for each of their columns in

$$L \, R_1 = \Phi \, \Delta_N \Sigma_u A_1^T + \Phi \, \Delta_S \Sigma_v A_2^T$$

$$L \, R_2 = \Phi \, \Delta_N \Sigma_u$$

3) Compute $[R_1 \; R_2]^T [R_1 \; R_2]$ and then $\tilde{\Omega}$ from (5.10).

Having computed $\tilde{\Omega}$, we use (4.16) to compute $\Omega = \mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}})$:

$$\Omega = \tilde{\Delta}_S^{-1} \tilde{\Omega} \tilde{\Delta}_S^{-T} = R_3 \tilde{\Delta}_S^{-T}$$

where $R_3 = \tilde{\Delta}_S^{-1} \tilde{\Omega}$ is obtained by solving

$$\tilde{\Delta}_S R_3 = \tilde{\Omega}$$

recursively for each column of $R_3$. Then $\Omega$ is obtained by solving

$$\Omega \tilde{\Delta}_S^T = R_3$$

recursively for each row of $\Omega$.

From the general considerations at the beginning of section 5, we need not compute all of $\Omega$. One could then avoid computing all of $R_3$ and $\tilde{\Omega}$. However, unlike section 5.1, here it does not seem possible with this approach to limit computing $\underset{\sim}{r}_i^T \underset{\sim}{r}_j$ (where $[R_1\ R_2] = [\underset{\sim}{r}_1, \ldots, \underset{\sim}{r}_n]$) to $i-j < k_1$ if $\mathrm{Cov}(S_{t-k} - \hat{S}_{t-k}, S_t - \hat{S}_t)$ is not of interest for $k > k_2$ for some small $k_1$ and $k_2$. We could achieve this reduction in computations by pre- and post-multiplying both terms in (5.10) by $\tilde{\Delta}_S^{-1}$ and $\tilde{\Delta}_S^{-T}$, respectively, yielding $([R_1\ R_2]\ \tilde{\Delta}_S^{-T})^T ([R_1\ R_2]\ \tilde{\Delta}_S^{-T})$ as the second term, rather than computing $\tilde{\Omega}$ first. This could be done with a partitioned inverse of $\tilde{\Delta}_S$, and in fact, other algebraic simplifications are then possible, which could give a more efficient algorithm. Unfortunately, pre- and post- multiplying by $\tilde{\Delta}_S^{-1}$ and $\tilde{\Delta}_S^{-T}$ accumulates entries down columns and across rows (e.g. if $\delta_S(B) = 1 - B$ then $\tilde{\Delta}_S^{-1}$ is lower triangular with 1 for each element). Thus, this approach results in subtracting one accumulating sum from another, which could easily result in an unstable algorithm.

### 5.3 Computing Results for Case III

The results of section 4.3 for the case where $S_t$ and $N_t$ are nonstationary, $\delta_S(B)$ and $\delta_N(B)$ have a common zero, Var $(\underset{\sim}{N}_*)$ is known, and $\underset{\sim}{N}_*$ is assumed independent of $\{u_t\}$ and $\{v_t\}$, may be computed using techniques developed in sections 5.1 and 5.2. One does need to recognize that $\Delta_N^* \sim \delta_N^*(B) = \delta_N(B)/\delta_c(B)$ and $\Delta_S^* \sim \delta_S^*(B) = \delta_S(B)/\delta_c(B)$.

# REFERENCES

Ansley, C. F. (1979), "An algorithm for the exact likelihood of a mixed autoregressive-moving average process," Biometrika, 66, 59-65.

Ansley, C. F. and Kohn R. (1985), "Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions," Annals of Statistics, 13, 1286-1316.

Bell, W. R. (1984), "Signal Extraction for Nonstationary Time Series," 12, 646-664.

Bell, W. R. and Hillmer, S. C. (1987a), "Initializing the Kalman Filter in the Nonstationary Case: With Application to Signal Extraction," SRD Research Report No. 87/33, Bureau of the Census.

_____(1987b), "Time Series Methods for Survey Estimation," SRD Research Report No. 87/20, Bureau of the Census.

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

Brockwell, P. J. and Davis, R. A. (1987), Time Series: Theory and Methods, New York: Springer-Verlag.

Burridge, P. and Wallis, K. F. (1985), "Calculating the Variance of Seasonally Adjusted Series," Journal of the American Statistical Association, 80, 541-552.

Carlin, J, (1987), "Seasonal Analysis of Economic Time Series," unpublished Ph.D. dissertation, Harvard University, Department of Statistics.

Cleveland, W. P. and Tiao, G. C. (1976), "Decomposition of Seasonal Time Series: A Model for the X-11 Program," Journal of the American Statistical Association, 71, 581-587.

Dent, W. (1977), "Computation of the Exact Likelihood Function of an ARIMA Process, " Journal of Statistical Computation and Simulation, 5, 193-206.

Dongarra, J. J., Moler, C. B., Bunch, J. R. and Stewart, G. W. (1979), Linpack User's Guide, Philadelphia: Society for Industrial and Applied Mathematics.

Gardner, G., Harvey, A. C. and Phillips, G.D.A. (1980), "An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering," Applied Statistics, 29, 311-322.

Hannan, E. J. (1970), Multiple Time Series, New York: Wiley.

_____ (1967), "Measurement of a Wandering Signal Amid Noise," Journal of Applied Probability, 4, 90-102.

Hillmer, S. C. (1985), "Measures of Variability for Model-Based Seasonal Adjustment Procedures," Journal of Business and Economic Statistics, 3, 60-68.

Hillmer, S. C. and Tiao, G. C. (1979), "Likelihood Function of Stationary Multiple Autoregressive Moving Average Models," Journal of the American Statistical Association, 74, 652-660.

Hotta, L. K. (1988), "Identification of UCARIMA Models," Journal of Time Series Analysis, to appear.

Jones, R. G. (1980), "Best Linear Unbiased Estimators for Repeated Surveys," Journal of the Royal Statistical Society, 42, 221-226.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series with Missing Observations," Technometrics, 22, 389-395.

Kitagawa, G. (1981), "A Nonstationary Time Series Model and Its Fitting by a Recursive Filter," Journal of Time Series Analysis, 2, 103-116.

Kohn, R. and Ansley, C. F. (1984), "A Note on Kalman Filtering for the Seasonal Moving Average Model," Biometrika, 71, 648-650.

_____ (1986), "Estimation, Prediction, and Interpolation for ARIMA Models With Missing Data," Journal of the American Statistical Association, 81, 751-761.

_____ (1987), "Signal Extraction for Finite Nonstationary Time Series," Biometrika, 74, 411-421.

Ljung, G. M. and Box, G.E.P. (1979), "The Likelihood Function of Stationary Autoregressive-Moving Average Models," Biometrika, 66, 265-270.

McLeod, I. (1975), "Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series," Applied Statistics, 24, 255-256.

_____ (1977), "Correction to Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series," Applied Statistics, 26, 194.

Mélard, G. (1984), "A Fast Algorithm for the Exact Likelihood of Autoregressive-Moving Average Models," Applied Statistics, 33, 104-114.

Newbold, P. (1974), "The Exact Likelihood Function for a Mixed Autoregressive-Moving Average Process," Biometrika, 61, 423-426.

Osborn, D. R. (1977), "Exact and Approximate Maximum Likelihood Estimators for Vector Moving Average Processes," Journal of the Royal Statistical Society, B 39, 114-118.

Otto, M. C., Bell, W. R. and Burman, J. P. (1987), "An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors, SRD Research Report No. 87/34, Bureau of the Census.

Pagan, A. (1975), "A Note on the Extraction of Components from Time Series," Econometrica, 43, 163-168.

Pearlman, J. G. (1980), "An Algorithm for the Exact Likelihood of a High-Order Autoregressive-Moving Average Process," Biometrika, 67, 232-233.

Phadke, M. S. and Kedem, G. (1978), "Computation of the Exact Likelihood Function of Multivariate Moving Average Models," Biometrika, 65, 511-519.

Pierce, D. A. (1979), "Signal Extraction Error in Nonstationary Time Series," Annals of Statistics, 7, 1303-1320.

Scott, A. J. and Smith, T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," Journal of the American Statistical Association, 69, 674-678.

Sobel, E. L. (1967), "Prediction of a Noise-distorted, Multivariate, Non-stationary Signal," Journal of Applied Probability, 4, 330-342.

Solo, V. (1986), "Topics in Advanced Time Series Analysis," in Lectures in Probability and Statistics, ed. G. del Pino and R. Rebolledo, Lecture Notes in Mathematics No. 1215, Springer-Verlag, pp. 165-328.

Tunnicliffe-Wilson, G. (1983), "The Estimation for Time Series Models, Part I. Yet Another Algorithm for the Exact Likelihood of ARMA Models," Technical Report No. 2528, Mathematics Research Center, University of Wisconsin-Madison.

Whittle, P. (1963), Prediction and Regulation by Linear Least Squares Methods, London: English Universities Press.

Wincek, M. A. and Reinsel, G. C. (1984), "An Exact Maximum Likelihood Estimation Procedure for Regression-ARMA Time Series Models with Possibly Nonconsecutive Data," Technical Report No. 749, Department of Statistics, University of Wisconsin-Madison.