BUREAU OF THE CENSUS
STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR–87/33

INITIALIZING THE KALMAN FILTER FOR NONSTATIONARY TIME SERIES MODELS

by

William Bell Statistical Research Division U.S. Bureau of the Census Washington, DC 20233	and	Steven Hillmer School of Business University of Kansas Lawrence, KS 66045

Report issued: October, 1987

Revised: February, 1991

INITIALIZING THE KALMAN FILTER FOR NONSTATIONARY TIME SERIES MODELS

by

William Bell
Statistical Research Division
U.S. Census Bureau

and

Steven Hillmer
School of Business
University of Kansas

February, 1991

Acknowledgment: We wish to acknowledge helpful comments of the referee, and helpful discussions with David F. Findley. Peter Zadrozny suggested the approach to general missing data problems in section 7. Marian Pugh assisted with the computations and graph for section 8. The research was based upon work supported by the National Science Foundation under grants SES 84-01460 and SES 86-09436, and by the Census Bureau under joint statistical agreement JSA 88-27. The research was partially conducted at the U.S. Bureau of the Census while the second author was a participant in the American Statistical Association/Census Bureau Research Program, which is supported by the Census Bureau and the NSF. Any opinions, findings, and conclusions expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation or the Census Bureau.

Abstract

The problem of initializing the Kalman filter for nonstationary time series models is considered. Ansley and Kohn (1985a) and Kohn and Ansley (1986) develop a "modified Kalman filter" for use with nonstationary models to produce estimates from what they call a "transformation approach". We show the same results can be obtained with a suitable initialization of the ordinary Kalman filter. Assuming there are d starting values for the nonstationary series, we initialize the Kalman filter using data through time d with the transformation approach estimate of the state vector and its associated error covariance matrix at time d. We give details of the initialization for ARIMA models, ARIMA component models, and dynamic linear models. We present an example to illustrate how the results may differ from results obtained under more naive initializations that have been suggested.

Keywords: modified Kalman filter, starting values, ARIMA model, ARIMA component model, dynamic linear model.

1. Introduction

The Kalman filter and variations of it have been widely advocated in recent years for time series filtering, prediction, interpolation, signal extraction, and likelihood evaluation. The algorithms require two things: (1) a known state-space model suitable for the problem, and (2) an estimate of the initial state vector and the variance of the error in this estimate. For stationary time series models the usual initialization uses the unconditional mean and variance of the initial state (see Akaike (1978) and Jones (1980).) For nonstationary time series models this approach is not available since unconditional means and covariances change over time and are typically only calculable given the initial state mean and covariance.

Various approaches have been taken to dealing with the difficulties of initializing the Kalman filter for nonstationary time series models. These include (1) letting the variance of the initial state be "large" (Harvey and Phillips 1979, Burridge and Wallis 1985); (2) using the information filter with the inverse of the initial state variance set to 0 (Kitagawa 1981); (3) augmenting the state vector for the differenced data with d observations (when differencing is of order d) and initializing at time d since the augmented part of the state vector is then known exactly (Harvey and Pierse 1984, also Harvey 1981 and Jones 1985); and (4) running the Kalman filter with initial state estimate and variance 0, and then computing an adjustment to yield results invariant to the mean and variance of the initial state (DeJong 1988). There are problems with implementation and justification for all these approaches. Ansley and Kohn (1985a), hereafter AK, point out that letting the initial variance be "large" is subject to numerical difficulties, and that the information filter cannot be used for all problems, such as an ARIMA (autoregressive-integrated-moving average) model of order (p, d, q) with $p + d < q + 1$. The Harvey and Pierse (1984) approach does not directly apply to component models. Also, their augmented state vector can be considerably larger than needed. (In the "airline model" of Box and Jenkins (1976), the dimension would be 27, much larger than the minimum dimension of 14.) DeJong's (1988) approach appears to require substantial computation, since it involves running the Kalman filter for an additional vector and square matrix of the same dimension as the state vector. While DeJong's (1988) approach has some justification, in many cases it goes too far (as do (1) and (2)), since we are not completely ignorant regarding assumptions about the initial state vector, only about the part due to the nonstationary starting values.

The most rigorous approach to dealing with the problems of initializing the Kalman filter in the nonstationary case is that of AK and Kohn and Ansley (1986). They present a "modified Kalman filter" that allows the variance of the part of the initial state due to the nonstationary starting values to be infinite. They provide justification for their approach by showing it produces the same results as a "transformation approach" that eliminates the effect of the nonstationary starting values. (They also note the Harvey and Pierse (1984) approach will give the same results in problems where it applies). Kohn and Ansley (1987) show that the transformation approach estimate has minimum mean squared error (MMSE) among all linear estimates that eliminate the effect of the starting values. The transformation approach estimate is appealing because the estimate with MMSE among all linear estimates will depend on assumptions about the starting values (see Bell (1984) and Result 2 in section 4 later), and it is these starting values about which we have no basis for assumptions in the nonstationary case. A drawback to the modified Kalman filter is the "modified" part; existing Kalman filter software cannot be used. It is also conceptually complex because, as pointed out by AK, in letting the variance of the starting values go to infinity one cannot interchange the filtering and limiting operations.

Objectives of the Paper

The primary objectives of this paper are to show how the ordinary Kalman filter can be initialized to yield the transformation approach estimates of AK and to provide specific details of our initialization for ARIMA models, ARIMA component models and dynamic linear models. If d starting values are required for the nonstationary series Y(t), then we initialize the Kalman filter with the transformation approach estimate of the state vector at time d using data through time d, and the variance of the error in this estimate. Direct computation of these quantities is not difficult. In general, our approach avoids the need to use the modified Kalman filter and avoids the need to do any recursions at times 1, ..., d.

Section 2 sets up the basic problem and notation, and section 3 reviews the transformation approach and gives our initialization of the Kalman filter in the general nonstationary case. Section 4 establishes invariance and optimality results that are of interest in their own right, and provides a simple proof that our initialization approach in fact yields the transformation approach estimates. Sections 5 and 6 give specifics for our initialization of the Kalman filter with ARIMA component models (that for ARIMA models follows as a special case) and a dynamic linear model, respectively. Most of the paper assumes that none of the first d data points are missing, though any pattern of missing data subsequent to time d is allowed. Section 7 discusses extension of our approach to the case when some of the first d data points are missing. While our approach has some limitations in this case, it should be kept in mind that many time series problems do not involve missing data, and still more do not involve missing data in the first d time periods. (Here d is the order of the "differencing" operator; with only first differencing (d=1) this is assured.) Section 8 gives an example to illustrate potential differences between our approach and more naive initializations as discussed above.

For simplicity in what follows we shall assume all random variables are normal with mean 0. Nonzero means are easily handled by subtracting them off. If we do not have normality we simply replace conditional expectations by linear projections, and optimality results refer only to linear estimators. We will deal only with univariate observations, Y(t), though the ideas extend easily to the case of Y(t) a vector.

2. Preliminaries

We will consider the usual state space model with the f x 1 state vector $\underset{\sim}{X}(t)$ and scalar observations Y(t):

$$\underset{\sim}{X}(t+1) = \underset{\sim}{F}(t)\underset{\sim}{X}(t) + \underset{\sim}{G}(t)\underset{\sim}{\varepsilon}(t) \tag{2.1}$$

$$Y(t) = \underset{\sim}{H}'(t)\underset{\sim}{X}(t) + \gamma(t). \tag{2.2}$$

We are concerned with situations where the state vector at each time point depends linearly upon a d x 1 vector of "starting values" $\underset{\sim}{\eta}$, along with elements of stationary time series (hereafter, stationary elements). Then, as in AK, the state vector may be written as

$$\underset{\sim}{X}(t) = \underset{\sim}{\Phi}(t)\underset{\sim}{\eta} + \underset{\sim}{\nu}(t) \tag{2.3}$$

where $\underset{\sim}{\Phi}(t)$ is an f x d (nonrandom) matrix and $\underset{\sim}{\nu}(t)$ an f x 1 random vector. Also

$$Y(t) = \underset{\sim}{A}'(t)\underset{\sim}{\eta} + \omega(t) \tag{2.4}$$

where $\underset{\sim}{A}'(t) = \underset{\sim}{H}'(t)\underset{\sim}{\Phi}(t)$, while $\underset{\sim}{\nu}(t)$ and $\omega(t)$ are linear combinations of stationary elements. For most of the paper we assume that the n values Y(1), ..., Y(d), $Y(t_{d+1})$, ..., $Y(t_n)$ are observed so that none of the first d values of Y(t) are missing. Collecting (2.4) for t=1, ..., t_n we write

$$\underset{\sim}{Y} = \underset{\sim}{A}\,\underset{\sim}{\eta} + \underset{\sim}{\omega} \tag{2.5}$$

where $\underset{\sim}{Y} = [Y(1), \ldots, Y(t_n)]'$, $\underset{\sim}{\omega} = [\omega(1), \ldots, \omega(t_n)]'$, and the n x d matrix $\underset{\sim}{A}$ has rows $\underset{\sim}{A}'(1), \ldots, \underset{\sim}{A}'(t_n)$.

Let the notation $\underset{\sim}{Y}_j^i$ for $i \leq j$ denote the vector of observations $[Y(t_i), \ldots, Y(t_j)]'$, in particular, $\underset{\sim}{Y}_d^1 = [Y(1), \ldots, Y(d)]'$ and $\underset{\sim}{\omega}_d^1 = [\omega(1), \ldots, \omega(d)]'$. We make the notational convention that terms with $\underset{\sim}{Y}_j^i$ or $\underset{\sim}{\omega}_j^i$, etc., are not present in the expressions where they appear if $i>j$. Terms with vectors or matrices with a zero dimension are also not present. Such "terms not present" can generally be taken as 0 (appropriately dimensioned), as should be clear where this occurs. These conventions allow the expressions given here to apply directly to various particular cases, including the stationary case.

3. The Transformation Approach and Initializing the Kalman Filter

AK define the transformation approach in general as follows. Let the nonsingular n x n transformation matrix $\underset{\sim}{J} = [\underset{\sim}{J}_1{}' \; \underset{\sim}{J}_2{}']'$ be such that $\underset{\sim}{J}_1\underset{\sim}{A}$ is a d x d nonsingular matrix and $\underset{\sim}{J}_2\underset{\sim}{A} = \underset{\sim}{0}$. Let $\underset{\sim}{Z}_1 \equiv \underset{\sim}{J}_1\underset{\sim}{Y} = \underset{\sim}{J}_1\underset{\sim}{A}\underset{\sim}{\eta} + \underset{\sim}{J}_1\underset{\sim}{\omega}$, and $\underset{\sim}{Z}_2 \equiv \underset{\sim}{J}_2\underset{\sim}{Y} = \underset{\sim}{J}_2\underset{\sim}{\omega}$ (note (2.5)). Let X be a random variable for which,

$$X = \underset{\sim}{\alpha}'\underset{\sim}{\eta} + \zeta$$

with ζ a linear combination of stationary elements and $\underset{\sim}{\alpha}$ a d x 1 vector in the space spanned by $\underset{\sim}{A}(1), \ldots, \underset{\sim}{A}(t_n)$. For any $\underset{\sim}{b}$ such that $\underset{\sim}{\alpha} = \underset{\sim}{A}'\underset{\sim}{b}$, we have $X - \underset{\sim}{b}'\underset{\sim}{Y} = (\underset{\sim}{\alpha}'\underset{\sim}{\eta} + \zeta) - \underset{\sim}{b}'(\underset{\sim}{A}\underset{\sim}{\eta} + \underset{\sim}{\omega}) = \zeta - \underset{\sim}{b}'\underset{\sim}{\omega}$. The transformation approach estimate of X using $\underset{\sim}{Y}$ that eliminates $\underset{\sim}{\eta}$ is defined as

$$\hat{X} = \underset{\sim}{b}'\underset{\sim}{Y} + E(X - \underset{\sim}{b}'\underset{\sim}{Y}|\underset{\sim}{Z}_2) = \underset{\sim}{b}'\underset{\sim}{Y} + E(\zeta - \underset{\sim}{b}'\underset{\sim}{\omega}|\underset{\sim}{Z}_2) \tag{3.1}$$

with $\mathrm{Var}(X - \hat{X}) = \mathrm{Var}(\zeta - \underset{\sim}{b}'\underset{\sim}{\omega}|\underset{\sim}{Z}_2)$.

These can be computed since the covariance structure of $(\underset{\sim}{\omega}, \zeta)$ is known. Kohn and Ansley (1986) observe that the result (3.1) is invariant to alternative choices of the transformation matrix $\underset{\sim}{J}$ and vector $\underset{\sim}{b}$ satisfying the given conditions.

We can easily apply these results to estimation of the state vector $\underset{\sim}{X}(t)$. Let $\underset{\sim}{A}_d$ be the d x d matrix composed of the first d rows of $\underset{\sim}{A}$. Assuming that the matrix $\underset{\sim}{A}_d$ is nonsingular (as will be true in the cases we consider here), a transformation satisfying our requirements is

$$\begin{bmatrix} \underset{\sim}{Z}_1 \\ \underset{\sim}{Z}_2 \end{bmatrix} = \begin{bmatrix} \underset{\sim}{A}_d^{-1} & \underset{\sim}{0} \\ -\underset{\sim}{A}'(t_{d+1})\underset{\sim}{A}_d^{-1} & \\ \vdots & \underset{\sim}{I} \\ -\underset{\sim}{A}'(t_n)\underset{\sim}{A}_d^{-1} & \end{bmatrix} \underset{\sim}{Y}$$

From (2.3) and (2.5), $\underset{\sim}{X}(t) - \underset{\sim}{\Phi}(t)[\underset{\sim}{A}_d^{-1} \;\; \underset{\sim}{0}]\underset{\sim}{Y}$ does not involve $\underset{\sim}{\eta}$. Here the f x n matrix $\underset{\sim}{\Phi}(t)[\underset{\sim}{A}_d^{-1} \;\; \underset{\sim}{0}]$ plays the role of $\underset{\sim}{b}$ in (3.1) (since we are estimating an f x 1 vector). The transformation approach estimate of the state $\underset{\sim}{X}(t)$ based upon the data $\underset{\sim}{Y}$ is then

$$\hat{\underset{\sim}{X}}(t|n) = \underset{\sim}{\Phi}(t)\underset{\sim}{A}_d^{-1}\underset{\sim}{Y}_d^1 + E[\underset{\sim}{\nu}(t) - \underset{\sim}{\Phi}(t)\underset{\sim}{A}_d^{-1}\underset{\sim}{\omega}_d^1|\underset{\sim}{Z}_2], \tag{3.2}$$

with $\quad \mathrm{Var}[\underset{\sim}{X}(t) - \hat{\underset{\sim}{X}}(t|n)] = \mathrm{Var}[\underset{\sim}{\nu}(t) - \underset{\sim}{\Phi}(t)\underset{\sim}{A}_d^{-1}\underset{\sim}{\omega}_d^1|\underset{\sim}{Z}_2].$

Notice that if d=0 the $\underset{\sim}{\Phi}(t)\underset{\sim}{A}_d^{-1}\underset{\sim}{Y}_d^1$ term in (3.2) is not present, and the results (3.2) reduce to the usual stationary case results, $\hat{\underset{\sim}{X}}(t|n) = E[\underset{\sim}{X}(t)|\underset{\sim}{Y}]$, and $\mathrm{Var}[\underset{\sim}{X}(t) - \hat{\underset{\sim}{X}}(t|n)] = \mathrm{Var}[\underset{\sim}{X}(t)|\underset{\sim}{Y}]$.

The only restriction on t or n above is $n \geq d$ (so that $\underset{\sim}{A}$ in (2.4) contains $\underset{\sim}{A}_d$). Our approach to initializing the Kalman filter is to use (3.2) with t=n=d. From (2.3) write $\underset{\sim}{X}(d) = \underset{\sim}{\Phi}\underset{\sim}{\eta} + \underset{\sim}{\nu}$ where for convenience in notation we suppress the dependence on d in $\underset{\sim}{\Phi}$ and $\underset{\sim}{\nu}$. Then using only the data $\underset{\sim}{Y}_d^1 = \underset{\sim}{A}_d\underset{\sim}{\eta} + \underset{\sim}{\omega}_d^1$, $\underset{\sim}{Z}_2$ is not present in (3.2) and we have

$$\hat{X}(d|d) = \Phi A_d^{-1} Y_d^1,$$
$$\mathrm{Var}[\hat{X}(d) - X(d|d)] = \mathrm{Var}(\nu - \Phi A_d^{-1} \omega_d^1). \tag{3.3}$$

Use of (3.3) applies the transformation approach of AK to the smallest set of observations (Y_d^1) that will allow removal of the dependence of the state vector on η. Once the estimate and covariance matrix in (3.3) are computed, the ordinary Kalman filter can be used for the remaining estimated state vectors and covariance matrices. In the next section we show that this approach in fact produces the transformation approach estimates. Notice that if there are any missing values subsequent to time d, the Kalman filter can handle them as in Jones (1980). In section 7 we discuss how the approach can be extended to the case of missing data in the first d observations.

4. Properties of the Estimates

The "starting values" η that are eliminated by the transformation approach are random variables used to "start-up" a stochastic difference equation such as an ARIMA model. Since there are multiple choices for starting values (some examples will be given shortly), this raises the question of whether the transformation approach results are the same for different choices of starting values η to eliminate. To consider this let η_2 be an alternative d x 1 random vector of starting values, such that

$$\eta_2 = \Gamma\eta + \xi \tag{4.1}$$

where Γ is a d x d nonsingular matrix and ξ is a d x 1 linear combination of stationary elements. Re-expressing Y and X using (4.1). as $Y = A\Gamma^{-1}\eta_2 +$ $(\omega - A\Gamma^{-1}\xi)$ and $X = \alpha'\Gamma^{-1}\eta_2 + (\zeta - \alpha'\Gamma^{-1}\xi)$, and using $b'A = \alpha'$, we easily see that $X - \underset{\sim}{b}'\underset{\sim}{Y}$ does not depend on η_2, in fact $X - \underset{\sim}{b}'\underset{\sim}{Y} = \zeta - \underset{\sim}{b}'\underset{\sim}{\omega}$ as before. So the transformation approach estimate of X eliminating η_2 is $\hat{X}_2 = \underset{\sim}{b}'\underset{\sim}{Y} + E(\zeta - \underset{\sim}{b}'\underset{\sim}{\omega}|\underset{\sim}{Z}_2)$ $= \hat{X}$. We thus have the following.

Result 1:

The transformation approach estimates (3.1) and (3.2) are invariant to alternative choices of starting values η as in (4.1).

Examples:

(1) For an ARIMA (p,d,q) model with observations beginning at t=1:

(a) Kohn and Ansley (1986) use $\eta = [Y(1-d), \ldots, Y(0)]'$ and Ansley and Kohn (1985b) use $\eta = [Y(0), \ldots, Y(1-d)]'$

(b) In section 5 we use $\eta = [Y(1), \ldots, Y(d)]'$.

(c) An obvious alternative is $\eta = [Y(j), \ldots, Y(j+d-1)]'$ for any j before, during, or after the observed data.

(2) For ARIMA component models (see section 5 for notation) with observations beginning at t=1:

(a) Bell (1984) uses $\eta = [\underset{\sim}{S}_*' \; \underset{\sim}{N}_*']'$ with $\underset{\sim}{S}_* = [S(1), \ldots, S(ds)]'$, $\underset{\sim}{N}_* = [N(1), \ldots, N(dn)]'$,

(b) Kohn and Ansley (1987) use $\underset{\sim}{S}_* = [S(1-ds), \ldots, S(0)]'$ and $\underset{\sim}{N}_* = [N(1-dn), \ldots, N(0)]'$.

(c) In section 5 we use $\underset{\sim}{S}_* = [S(d-ds+1), \ldots, S(d)]'$ and $\underset{\sim}{N}_* = [N(d-dn+1), \ldots, N(d)]'$.

(d) If there are no common roots in $\delta_s(B)$ and $\delta_n(B)$, another choice is $\eta = [Y(1), \ldots, Y(d)]'$ (see (5.11)).

That these alternative choices of starting values satisfy (4.1) may be seen from results given in section 5.

The next result concerns optimality of the transformation approach estimate $\hat{X}$ of (3.1) (or $\hat{\underset{\sim}{X}}(t|n)$ of (3.2)). Consider the following assumption.

<u>Assumption A</u>: The vector $\underset{\sim}{Y}_d^1$ is independent of $\underset{\sim}{\omega}$ and ζ. (For estimating $\underset{\sim}{X}(t)$ replace ζ by $\underset{\sim}{\nu}(t)$.)

We can easily prove that the transformation approach estimate is MMSE among <u>all</u> linear estimates under Assumption A, and then use this result to provide a simple proof of a result of Kohn and Ansley (1987) that the transformation approach estimate is always MMSE among those linear estimates that eliminate the effect of starting values.

<u>Result 2:</u>

Assume we have the state space model (2.1) and (2.2) with all of Y(1),..., Y(d) observed. Then (a) Under Assumption A, AK's transformation approach yields the MMSE linear estimate $\hat{X} = E(X|\underset{\sim}{Y})$ (or $\hat{\underset{\sim}{X}}(t|n) = E[\underset{\sim}{X}(t)|\underset{\sim}{Y}]$). (b) (Kohn and Ansley 1987, Theorem 2.1) Let $\tilde{X}$ be any linear estimator of X using $\underset{\sim}{Y}$ such that $X-\tilde{X}$ does not depend on $\underset{\sim}{\eta}$. Then $\text{Var}(X-\hat{X}) \leq \text{Var}(X-\tilde{X})$ with equality holding if and only if $\tilde{X} = \hat{X}$ almost surely. (An analogous result holds for $\hat{\underset{\sim}{X}}(t|n)$.)

<u>Proof</u>: To prove (a) notice from $X - \underset{\sim}{b}'\underset{\sim}{Y} = \zeta - \underset{\sim}{b}'\underset{\sim}{\omega}$ and (3.1) it follows that $X - \hat{X} = \zeta - \underset{\sim}{b}'\underset{\sim}{\omega} - E(\zeta - \underset{\sim}{b}'\underset{\sim}{\omega}|\underset{\sim}{Z}_2)$. ζ, $\underset{\sim}{\omega}$, and $\underset{\sim}{Z}_2 = \underset{\sim}{J}_2\underset{\sim}{\omega}$ are independent of $\underset{\sim}{Z}_1 = \underset{\sim}{A}_d^{-1}\underset{\sim}{Y}_d^1$ under Assumption A, so $X - \hat{X}$ is independent of $\underset{\sim}{Z}_1$. But $\zeta - \underset{\sim}{b}'\underset{\sim}{\omega} - E(\zeta - \underset{\sim}{b}'\underset{\sim}{\omega}|\underset{\sim}{Z}_2)$ is orthogonal to, and thus independent of $\underset{\sim}{Z}_2$. Since $\underset{\sim}{Y} = \underset{\sim}{J}^{-1}[\underset{\sim}{Z}_1{}'\ \underset{\sim}{Z}_2{}']'$ we have $\underset{\sim}{Y}$ independent of $X - \hat{X}$ proving $\hat{X} = E(X|\underset{\sim}{Y})$. Similarly $\hat{\underset{\sim}{X}}(t|n) = E(\hat{\underset{\sim}{X}}(t)|\underset{\sim}{Y})$.

To prove (b) notice that it holds under Assumption A since then $\hat{X}$ is MMSE. But since $X - \hat{X}$ and $X - \tilde{X}$ depend only on $\underset{\sim}{\omega}$ and ζ (not $\underset{\sim}{\eta}$), and Assumption A does not deal with the covariance structure of ($\underset{\sim}{\omega}$, ζ), result (b) holds in general. It holds similarly for $\underset{\sim}{X}(t|n)$ in the sense that $\text{Var}(\underset{\sim}{X}(t) - \tilde{\underset{\sim}{X}}(t|n)) - \text{Var}(X(t) - \hat{\underset{\sim}{X}}(t|n))$ is positive semi-definite for suitable $\tilde{\underset{\sim}{X}}(t|n)$. QED

Result 2(a) tells us when the transformation approach estimate is optimal, but says nothing about the suitability of Assumption A in any particular application. (Bell (1984) discusses consequences of Assumption A and an alternative assumption in the signal extraction problem with the entire sequence Y(t), t = 0, ± 1, ± 2, ... available.) Result 2(b) is appealing because there is typically no basis for explicit assumptions about the starting values η. In such cases one cannot hope to obtain optimal estimates and it is sensible to consider sub-optimal estimates whose error covariances are known (the class $\tilde{X}$), rather than guess at assumptions for the starting values in an effort an effort to achieve the optimal estimates. Result 2(b) establishes that $\hat{X}$ is optimal within this restricted class of estimates.

We can use Result 2 to establish the equivalence of the transformation approach and our initialization coupled with use of the ordinary Kalman filter. By Result 2(a), under Assumption A, $\hat{\underset{\sim}{X}}(d|d)$ in (3.3) is the optimal estimate, $E[\underset{\sim}{X}(d)|\underset{\sim}{Y}_d^1]$. Subsequent application of a Kalman filter/smoother will yield optimal estimators and their error covariances for every successive state. Now since both the transformation approach and the Kalman filter/smoother initialized by (3.3) produce MMSE estimates under Assumption A, they must coincide under that assumption. But since neither of these approaches uses assumptions about η in deriving the estimators, they must produce the same results regardless of what is true of the starting values η. We thus have the following.

Result 3

If we have the state space model (2.1) and (2.2) with all of Y(1),..., Y(d) observed, then the Kalman filter initialized by (3.3) yields the transformation approach estimates of $\underset{\sim}{X}(t)$ using $\underset{\sim}{Y}_t^1$ for $t \geq d$. Applying a Kalman smoother yields the transformation approach estimates of $\underset{\sim}{X}(t)$ using $\underset{\sim}{Y}$.

5. Initialization of an ARIMA Components Model

In this section we show how to apply the initialization ideas of Section 3 in a signal extraction framework where the component models are of the ARIMA form. In particular, suppose the observed series Y(t) satisfies

$$Y(t) = S(t) + N(t). \tag{5.1}$$

The model for the signal component is given by $\delta_s(B)S(t) = U(t)$ and $\phi_s(B)U(t) = \theta_s(B)b(t)$ where $\delta_s(B) = 1 - \delta_{s1}B - \cdots - \delta_{s,ds}B^{ds}$ is a polynomial in the backshift operator B ($BS(t) = S(t-1)$) of degree ds whose zeros are on the unit circle, $\phi_s(B) = 1 - \phi_{s1}B - \cdots - \phi_{s,ps}B^{ps}$ is a polynomial in B of degree ps whose zeros are outside the unit circle, $\theta_s(B) = 1 - \theta_{s1}B - \cdots - \theta_{s,qs}B^{qs}$ is a polynomial in B of degree qs whose zeros are on or outside the unit circle, and b(t) is white noise with variance σ_b^2. Similarly, let the model for the noise component be given by $\delta_n(B)N(t) = V(t)$ and $\phi_n(B)V(t) = \theta_n(B)c(t)$. We assume the time series U(t) and V(t) are uncorrelated with each other. The polynomials $\delta_n(B)$, $\phi_n(B)$, and $\theta_n(B)$ are of degrees dn, pn, and qn respectively, and are defined in an analogous way to the corresponding signal component polynomials. We assume that $\delta_s(B)$ and $\delta_n(B)$ have no common zeros. Let $\tilde{\phi}_s(B) = 1 - \tilde{\phi}_{s1}B - \cdots - \tilde{\phi}_{s,ps+ds}B^{ps+ds} = \delta_s(B)\phi_s(B)$, a polynomial of degree ps+ds, and similarly $\tilde{\phi}_n(B) = \delta_n(B)\phi_n(B)$, a polynomial of degree pn+dn.

Following Kohn and Ansley (1986) a state space representation for S(t) of minimal degree fs = max(ps+ds,qs+1) is

$$\begin{aligned} \underset{\sim}{X}_s(t+1) &= \underset{\sim}{F}_s \underset{\sim}{X}_s(t) + \underset{\sim}{G}_s b(t+1) \\ S(t+1) &= \underset{\sim}{H}_s' \underset{\sim}{X}_s(t+1) \end{aligned} \tag{5.2}$$

In (5.2) the fs x 1 state vector has the components $X_{s1}(t) = S(t)$ and

$$X_{si}(t) = \sum_{j=i}^{ps+ds} \tilde{\phi}_{sj} S(t-1+i-j) - \sum_{j=i-1}^{qs} \theta_{sj} b(t-1+i-j) \text{ for } i=2, \ldots, fs.$$

The fs x 1 vectors $\underset{\sim}{H}_s = (1, 0, \ldots 0)'$ and $\underset{\sim}{G} = (1, -\theta_{s1}, \cdots -\theta_{s,fs-1})'$ and the fs x fs matrix

$$\underset{\sim}{F}_s = \begin{bmatrix} \tilde{\phi}_{s,1} & 1 & \cdot & & \\ \vdots & & & \cdot & \\ & & & & \cdot \\ \tilde{\phi}_{s,fs-1} & 0 & 0 & \cdots & 1 \\ \tilde{\phi}_{s,fs} & 0 & 0 & \cdots & 0 \end{bmatrix}$$

where $\tilde{\phi}_{s,j} = 0$ for $j > ps+ds$ and $\theta_{s,j} = 0$ for $j > qs$. In an analogous manner a state space representation for N(t) of degree fn = max(pn+dn, qn+1) is

$$\begin{aligned} \underset{\sim}{X}_n(t+1) &= \underset{\sim}{F}_n \underset{\sim}{X}_n(t) + \underset{\sim}{G}_n c(t+1) \\ N(t+1) &= \underset{\sim}{H}_n' \underset{\sim}{X}_n(t+1) \end{aligned} \qquad (5.3)$$

From (5.1), (5.2), and (5.3) it follows that a state space representation for Y(t) is

$$\begin{aligned} \underset{\sim}{X}(t+1) &= \underset{\sim}{F}\underset{\sim}{X}(t) + \underset{\sim}{G}\underset{\sim}{\varepsilon}(t+1) \\ Y(t+1) &= \underset{\sim}{H}'\underset{\sim}{X}(t+1) \end{aligned} \qquad (5.4)$$

where $\underset{\sim}{X}'(t) = [\underset{\sim}{X}_s'(t), \underset{\sim}{X}_n'(t)]$, $\underset{\sim}{\varepsilon}'(t) = [b(t), c(t)]$, $\underset{\sim}{H}' = [\underset{\sim}{H}_s', \underset{\sim}{H}_n']$, $\underset{\sim}{F} = \text{diag}\,(\underset{\sim}{F}_s, \underset{\sim}{F}_n)$, and $\underset{\sim}{G} = \text{diag}(\underset{\sim}{G}_s, \underset{\sim}{G}_n)$ where diag $(\underset{\sim}{A}, \underset{\sim}{B})$ denotes a block diagonal matrix of appropriate dimensions. From the way the state vectors are defined we can write

$$\begin{aligned} \underset{\sim}{X}_s(t) &= \tilde{\underset{\sim}{\Phi}}_s \underset{\sim}{S}_t^{t+1-rs} + \underset{\sim}{\theta}_s \underset{\sim}{b}_t^{t+1-qs} \\ \underset{\sim}{X}_n(t) &= \tilde{\underset{\sim}{\Phi}}_n \underset{\sim}{N}_t^{t+1-rn} + \underset{\sim}{\theta}_n \underset{\sim}{c}_t^{t+1-qn} \end{aligned} \qquad (5.5)$$

where rs=max(ps+ds,1), rn=max(pn+dn,1), and the fs x rs matrix $\tilde{\underset{\sim}{\Phi}}_s$ and the fs x qs matrix $\underset{\sim}{\theta}_s$ are given by

$$\tilde{\underset{\sim}{\Phi}}_s = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ \tilde{\phi}_{s,rs} & \tilde{\phi}_{s,rs-1} & \cdots & \tilde{\phi}_{s,2} & 0 \\ 0 & \tilde{\phi}_{s,rs} & \cdots & \tilde{\phi}_{s,3} & 0 \\ \vdots & & & & \\ 0 & 0 & \cdots & \tilde{\phi}_{s,rs} & 0 \\ & \underset{\sim}{0}_{(fs-rs)\times rs} & & & \end{bmatrix} \quad \underset{\sim}{\theta}_s = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ -\theta_{s,qs} & -\theta_{s,qs-1} & \cdots & -\theta_{s,1} \\ 0 & -\theta_{s,qs} & \cdots & -\theta_{s,2} \\ \vdots & \vdots & & \\ 0 & 0 & & -\theta_{s,qs} \\ & \underset{\sim}{0}_{(fs-qs-1)\times qs} & & \end{bmatrix}. \qquad (5.6)$$

The fn x rn matrix $\tilde{\underset{\sim}{\Phi}}_n$ and the fn x qn matrix $\underset{\sim}{\theta}_n$ are defined analogously. From

(5.5) then

$$\underset{\sim}{X}(t) = \tilde{\underset{\sim}{\Phi}}_1 \begin{bmatrix} \underset{\sim}{S}_t^{t+1-rs} \\ \underset{\sim}{N}_t^{t+1-rn} \end{bmatrix} + \underset{\sim}{\theta}_1 \begin{bmatrix} \underset{\sim}{b}_t^{t+1-qs} \\ \underset{\sim}{c}_t^{t+1-qn} \end{bmatrix} \tag{5.7}$$

where $\tilde{\underset{\sim}{\Phi}}_1 = \mathrm{diag}\,(\tilde{\underset{\sim}{\Phi}}_s, \tilde{\underset{\sim}{\Phi}}_n)$ and $\underset{\sim}{\theta}_1 = \mathrm{diag}\,(\underset{\sim}{\theta}_s, \underset{\sim}{\theta}_n)$.

Given the models for S(t) and N(t) there are ds starting values for S(t) and dn starting values for N(t) needed. We shall take $\underset{\sim}{S}_* = [S(d+1-ds),\ldots,S(d)]'$ where d = ds + dn as the starting values for S(t), and $\underset{\sim}{N}_* = [N(d+1-dn),\ldots,N(d)]'$ as the starting values for N(t). Given $\delta_s(B)$, following Bell(1984) we define the quantities

$$A_{j,t}^s = \begin{cases} 1 & j=t \quad t=1,\ \ldots,\ ds \\ 0 & j\neq t \quad t=1,\ \ldots,\ ds \\ \delta_{s,1}\, A_{j,t-1}^s + \ldots + \delta_{s,ds}\, A_{j,t-ds}^s & \text{if } t > ds \\ \delta_{s,1}\, A_{j,t+1}^s + \ldots + \delta_{s,ds}\, A_{j,t+ds}^s & \text{if } t \leq 0\ . \end{cases}$$

Let $\underset{\sim}{A}_s'(t) = [A_{1,t}^s,\ldots,A_{ds,t}^s]$, and define the values ξ_j^s by equating coefficients of B^j in $(\xi_0^s + \xi_1^s B + \xi_2^s B^2 + \ldots)\delta_s(B) = 1$. In the same way given $\delta_n(B)$ we define $\underset{\sim}{A}_n'(t)$ and ξ_j^n. Then we have that

$$S(t) = \begin{cases} \underset{\sim}{A}_s'(t-dn)\underset{\sim}{S}_* & \text{if } t=dn+1,\ldots,d \\ \underset{\sim}{A}_s'(t-dn)\underset{\sim}{S}_* + \sum_{i=0}^{t-d-1} \xi_i^s U(t-i) & \text{if } t > d \\ \underset{\sim}{A}_s'(t-dn)\underset{\sim}{S}_* + (-1)^{\rho s} \sum_{i=0}^{dn-t} \xi_i^s U(t+ds+i) & \text{if } t \leq dn \end{cases} \tag{5.8}$$

where ρs is the number of times (1-B) appears in $\delta_s(B)$. An analogous representation holds for N(t).

Making use of (5.8) and the analog for N(t), one can show that

$$\begin{bmatrix} \underset{\sim}{S}_d^{d+1-rs} \\ \underset{\sim}{N}_d^{d+1-rn} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{A}_{rs}^s & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{A}_{rn}^n \end{bmatrix} \begin{bmatrix} \underset{\sim}{S}_* \\ \underset{\sim}{N}_* \end{bmatrix} + \begin{bmatrix} \underset{\sim}{C}_{rs}^s & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{C}_{rn}^n \end{bmatrix} \begin{bmatrix} \underset{\sim}{U}_d^{ms} \\ \underset{\sim}{V}_d^{mn} \end{bmatrix} \tag{5.9}$$

where the rs x ds matrix $\underset{\sim}{A}^s_{rs}$ and the rs x (rs−ds) matrix $\underset{\sim}{C}^s_{rs}$ are

$$\underset{\sim}{A}^s_{rs} = \begin{bmatrix} \underset{\sim}{A}'_s(1-p) \\ \vdots \\ \underset{\sim}{A}'_s(0) \\ \underset{\sim}{I}_{ds} \end{bmatrix}, \quad \underset{\sim}{C}^s_{rs} = (-1)^{\rho s} \begin{bmatrix} \xi^s_0 & \xi^s_1 & \cdots & \xi^s_{rs-ds-1} \\ 0 & \xi^s_0 & \cdots & \xi^s_{rs-ds-2} \\ \vdots & & & \\ 0 & 0 & \cdots & \xi^s_0 \\ & \underset{\sim}{0}_{ds \times (rs-ds)} & & \end{bmatrix} \tag{5.10}$$

and ms = d+1−rs+ds. The rn x dn matrix $\underset{\sim}{A}^n_{rn}$ and the rn x (rn−dn) matrix $\underset{\sim}{C}^n_{rn}$ are defined analogously, and mn = d+1−rn+dn. Furthermore, from (5.1) and using (5.8) and the analog for N(t), we write $\underset{\sim}{Y}^1_d = \underset{\sim}{S}^1_d + \underset{\sim}{N}^1_d$ as

$$\underset{\sim}{Y}^1_d = \underset{\sim}{A}_d \begin{bmatrix} \underset{\sim}{S}_* \\ \\ \underset{\sim}{N}_* \end{bmatrix} + \underset{\sim}{C}^s_d \underset{\sim}{U}^{ds+1}_d + \underset{\sim}{C}^n_d \underset{\sim}{V}^{dn+1}_d \tag{5.11}$$

where the d x d matrix $\underset{\sim}{A}_d = [\underset{\sim}{A}^s_d \ \underset{\sim}{A}^n_d]$, with the d x ds matrix $\underset{\sim}{A}^s_d$ and the d x (d-ds) matrix $\underset{\sim}{C}^s_d$ given by

$$\underset{\sim}{A}^s_d = \begin{bmatrix} \underset{\sim}{A}'_s(1-dn) \\ \vdots \\ \underset{\sim}{A}'_s(0) \\ \underset{\sim}{I}_{ds} \end{bmatrix} \quad \underset{\sim}{C}^s_d = (-1)^{\rho s} \begin{bmatrix} \xi^s_0 & \xi^s_1 & \cdots & \xi^s_{dn-1} \\ 0 & \xi^s_0 & \cdots & \xi^s_{dn-2} \\ \vdots & & & \\ 0 & 0 & \cdots & \xi^2_0 \\ & \underset{\sim}{0}_{ds \times d-ds} & & \end{bmatrix}.$$

The d x dn matrix $\underset{\sim}{A}^n_d$ and the d x (d−dn) matrix $\underset{\sim}{C}^n_d$ are similarly defined. It can be shown by the same arguments as in Bell(1982) that when $\delta_s(B)$ and $\delta_s(B)$ have no common zeros $\underset{\sim}{A}_d$ will be nonsingular.

We will initialize the Kalman filter at time d so that making use of (5.7) when t=d and the relationship in (5.9) we have that

$$\underset{\sim}{X}(d) = \underset{\sim}{\Phi}\begin{bmatrix} \underset{\sim}{S}_* \\ \\ \underset{\sim}{N}_* \end{bmatrix} + \tilde{\underset{\sim}{\Phi}}_1 \begin{bmatrix} \underset{\sim}{C}_{rs}^{s} & \underset{\sim}{0} \\ \\ \underset{\sim}{0} & \underset{\sim}{C}_{rn}^{n} \end{bmatrix} \begin{bmatrix} \underset{\sim}{U}_d^{ms} \\ \\ \underset{\sim}{V}_d^{mn} \end{bmatrix} + \underset{\sim}{\theta}_1 \begin{bmatrix} \underset{\sim}{b}_d^{d+1-qs} \\ \\ \underset{\sim}{c}_d^{d+1-qn} \end{bmatrix} \tag{5.12}$$

where $\underset{\sim}{\Phi} = \underset{\sim}{\Phi}_1 \operatorname{diag}(\underset{\sim}{A}_{rs}^{s}, \underset{\sim}{A}_{rn}^{n}) = \operatorname{diag}(\tilde{\underset{\sim}{\Phi}}_s \underset{\sim}{A}_{rs}^{s}, \tilde{\underset{\sim}{\Phi}}_n \underset{\sim}{A}_{rn}^{n})$. By comparing (5.11) and (5.12) we can estimate $\underset{\sim}{X}(d)$ by

$$\hat{\underset{\sim}{X}}(d|d) = \underset{\sim}{\Phi}\underset{\sim}{A}_d^{-1}\underset{\sim}{Y}_d^{1}. \tag{5.13}$$

Let ks = min(ms,ds+1) and kn = min(mn,dn+1), and define the rs x (d+1-ks) matrix $\tilde{\underset{\sim}{C}}_s^{s} = [\underset{\sim}{0}_{rs \times (ms-ks)}, \underset{\sim}{C}_{rs}^{s}]$, the d x (d+1-ks) matrix $\tilde{\underset{\sim}{C}}_d^{s} = [\underset{\sim}{0}_{d \times (ds+1-ks)}, \underset{\sim}{C}_d^{s}]$, the rn x (d+1-kn) matrix $\tilde{\underset{\sim}{C}}_{rn}^{n} = [\underset{\sim}{0}_{rn \times (mn-kn)}, \underset{\sim}{C}_{rn}^{n}]$ and the d x (d+1-kn) matrix $\tilde{\underset{\sim}{C}}_d^{n} = [\underset{\sim}{0}_{d \times (dn+1-kn)}, \underset{\sim}{C}_d^{n}]$, then it can be shown that

$$\begin{aligned} \operatorname{Var}[\underset{\sim}{X}(d) - \hat{\underset{\sim}{X}}(d|d)] &= \underset{\sim}{M} \operatorname{diag}[\operatorname{Var}(\underset{\sim}{U}_d^{ks}), \operatorname{Var}(\underset{\sim}{V}_d^{kn})]\, \underset{\sim}{M}' \\ &\quad + \operatorname{diag}[\sigma_b^2\, \underset{\sim}{\theta}_s \underset{\sim}{\theta}_s', \sigma_c^2 \underset{\sim}{\theta}_n \underset{\sim}{\theta}_n'] + \underset{\sim}{Q} + \underset{\sim}{Q}' \end{aligned} \tag{5.14}$$

where $\underset{\sim}{M} = \operatorname{diag}(\tilde{\underset{\sim}{\Phi}}_s, \tilde{\underset{\sim}{\Phi}}_n)\{\operatorname{diag}(\tilde{\underset{\sim}{C}}_{rs}^{s}, \tilde{\underset{\sim}{C}}_{rn}^{n}) - \operatorname{diag}(\underset{\sim}{A}_{rs}^{s}, \underset{\sim}{A}_{rn}^{n})\underset{\sim}{A}_d^{-1} [\tilde{\underset{\sim}{C}}_d^{s}\ \tilde{\underset{\sim}{C}}_d^{n}]\}$ and

$$\underset{\sim}{Q} = \underset{\sim}{M} \operatorname{diag}[\operatorname{Cov}(\underset{\sim}{U}_d^{ks}, \underset{\sim}{b}_d^{d+1-qs})\underset{\sim}{\theta}_s', \operatorname{Cov}(\underset{\sim}{U}_d^{kn}, \underset{\sim}{c}_d^{d+1-qn})\underset{\sim}{\theta}_n']$$

In (5.14) $\operatorname{Var}(\underset{\sim}{U}_d^{ks})$ is a covariance matrix from the ARMA(ps,qs) model for U(t), and $\operatorname{Var}(\underset{\sim}{V}_d^{kn})$ is a covariance matrix from the ARMA(pn,qn) model for V(t). McLeod (1975, 1977) shows how to compute such covariances. $\operatorname{Cov}(\underset{\sim}{U}_d^{ks}, \underset{\sim}{b}_d^{d+1-qs})$ contains the elements

$$\mathrm{Cov}(U(t),b(j)) = \begin{cases} 0 & \text{if } j > t \\ \sigma_b^2 \psi_{t-j}^s & \text{if } j \le t \end{cases}$$

with $(\psi_0^s + \psi_1^s B + \ldots) = \psi^s(B) = \theta_s(B)\phi_s^{-1}(B)$, and $\mathrm{Cov}(\underset{\sim}{V}_d^{kn}, \underset{\sim}{c}_d^{d+1-qn})$ contains the similarly defined elements Cov(V(t),c(j)). We thus have our initialization for ARIMA component models.

<u>Result 4 (Initialization of the Kalman Filter/Smoother for ARIMA Component Models and Signal Extraction)</u>:

Let Y(t) = S(t) + N(t) where the components S(t) and N(t) satisfy the ARIMA models following (5.1), with state space representations for S(t), N(t), and Y(t) in (5.2), (5.3), and (5.4). Then the same results as the transformation approach (and AK's modified Kalman filter/smoother) are obtained by initializing the ordinary Kalman filter/smoother at time d with $\hat{\underset{\sim}{X}}(d|d) = \underset{\sim}{\Phi}\underset{\sim}{A}_d^{-1}\underset{\sim}{Y}_d^1$ (see (5.13) and definitions preceding) and error covariance matrix given by (5.14).

<u>Special Case: A Simple ARIMA Model</u>

The preceding results easily specialize to provide an initialization of the Kalman filter for a simple ARIMA model, $\phi(B)\delta(B)Y(t) = \theta(B)a(t)$ (to eliminate $\underset{\sim}{\eta}=\underset{\sim}{Y}_d^1$). Drop N(t) from consideration so $Y(t) \equiv S(t)$, let $\delta(B)Y(t) = W(t) \equiv U(t)$, and drop the "s" indication from all relevant quantities since they now refer to Y(t). Now $\delta(B) = \delta_s(B)$, in which case $\underset{\sim}{A}_d = \underset{\sim}{I}$. Then the estimate (5.13) becomes

$$\hat{\underset{\sim}{X}}(d|d) = \tilde{\underset{\sim}{\Phi}}_1\underset{\sim}{A}_r\underset{\sim}{Y}_d^1 \tag{5.15}$$

where $\tilde{\underset{\sim}{\Phi}}_1$ corresponds to $\tilde{\phi}(B) = \phi(B)\delta(B)$ as in (5.6), and $\underset{\sim}{A}_r = [\underset{\sim}{A}(1-p), \cdots \underset{\sim}{A}(0), \underset{\sim}{I}_d]'$ with the A(t) defined from $\delta(B)$ as in (5.10). Also (5.11) reduces to $\underset{\sim}{Y}_d = \underset{\sim}{S}_d$ and $\tilde{\underset{\sim}{\Phi}}_1\underset{\sim}{C}_r$ corresponds to $\underset{\sim}{M}$, so (5.14) reduces to

$$\text{Var}[\underset{\sim}{X}(d) - \hat{\underset{\sim}{X}}(d|d)] = \tilde{\underset{\sim}{\Phi}}_1 \underset{\sim}{C}\,\text{Var}[\underset{\sim}{W}_d^m]\underset{\sim}{C}'\tilde{\underset{\sim}{\Phi}}_1' + \sigma_a^2 \underset{\sim}{\theta}\underset{\sim}{\theta}'$$

$$+ \tilde{\underset{\sim}{\Phi}}_1 \underset{\sim}{C}\,\text{Cov}(\underset{\sim}{W}_d^m, \underset{\sim}{a}_d^{d+1-q})\underset{\sim}{\theta}' + \underset{\sim}{\theta}\,\text{Cov}(\underset{\sim}{a}_d^{d+1-q}, \underset{\sim}{W}_d^m)\underset{\sim}{C}'\tilde{\underset{\sim}{\Phi}}_1' \qquad (5.16)$$

where $\underset{\sim}{\theta}$ corresponds to $\theta(B)$ as in (5.6), $\underset{\sim}{C}$ is obtained from $\delta(B)$ as in (5.10), $r=\max(p+d,1)$, and $m = 2d+1-r$.

The covariance initialization (5.16) is, apart from slight differences in notation and approach, the same as the initialization of the stationary part of the modified Kalman filter at time 0 given in Ansley and Kohn (1985b, section 4). They give an expression (Ansley and Kohn 1985b, (2.11)) for $\underset{\sim}{X}(0)$ that is analogous to (5.12) for one component, but with everything translated back d units in time, including defining $\underset{\sim}{\eta}$ = [Y(0)), ..., Y(1−d)]′. By initializing at time d, we avoid the need to do any recursions at times t = 1, ...,d, and avoid the need to use the modified Kalman filter. If there are no missing observations, then the transformation approach results for likelihood evaluation can be obtained by simply differencing the data and working with the stationary ARIMA model for W(t) = $\delta(B)Y(t)$ (note AK, p. 1290).

In using (5.15) and (5.16), we assume the stationary distribution for $\underset{\sim}{W}_d^m$ This is essentially assuming that the time series W(t) started in the remote past. However, we may wish to investigate the possibility that one or more zeros of $\phi(B)$ may be on or inside the unit circle. Enforcing the stationary assumption on $\phi(B)$ presents theoretical and computational difficulties in this case. When we are not willing to make the stationarity assumption for $\underset{\sim}{W}(t)$, we can incorporate $\phi(B)$ into $\delta(B)$ and use (5.15) and (5.16) with $\tilde{\phi}(B)$ replacing $\delta(B)$ and $\phi(B)=1$. Here we need $r=p+d$ starting values for Y(t), say $\underset{\sim}{\eta} = \underset{\sim}{Y}_r^1$. We now initialize the Kalman filter at time t=r, and the results (5.15) and (5.16) simplify to $\hat{\underset{\sim}{X}}(r|r) = \tilde{\underset{\sim}{\Phi}}_1\underset{\sim}{Y}_r^1$ and

$\text{Var}[\underset{\sim}{X}(r) - \hat{\underset{\sim}{X}}(r|r)] = \sigma_a^2 \underset{\sim}{\theta}\underset{\sim}{\theta}'$. A similar approach can be used with the ARIMA components model to avoid the stationarity assumption on U(t) or V(t) or both.

6. Initialization of a Dynamic Linear Model

In this section we illustrate how the initialization ideas of section 3 apply in the framework of a regression model with time varying parameters. In particular, suppose for t = 1, ..., n that

$$Y(t) = \underset{\sim}{H}'(t)\, \underset{\sim}{\beta}(t) + \gamma(t) \tag{6.1}$$

$$\underset{\sim}{\beta}(t) = \underset{\sim}{F}\underset{\sim}{\beta}(t-1) + \underset{\sim}{\varepsilon}(t). \tag{6.2}$$

Here Y(t) is an observed dependent variable, $\underset{\sim}{H}'(t)$ is a known 1 x d vector of independent variables, $\underset{\sim}{\beta}(t)$ is a d x 1 vector of time dependent parameters (the unknown state vector), $\underset{\sim}{F}$ is a known d x d matrix, the $\gamma(t)$ are iid normal random variables with mean zero and variance σ_γ^2, and the $\underset{\sim}{\varepsilon}(t)$ are iid d x 1 normal random vectors with mean $\underset{\sim}{0}$ and $\text{Var}[\underset{\sim}{\varepsilon}(t)] = \underset{\sim}{\Sigma}_\varepsilon$. Equations (6.1) and (6.2) are a particular case of the dynamic linear model of Harrison and Stevens (1976), and the time varying parameter regression model of Machak, Spivey, and Wrobleski (1985). Following the approach of section 3 we let the starting values be

$$\underset{\sim}{\eta} = \underset{\sim}{\beta}(d). \tag{6.3}$$

Then it follows directly from (6.1) to (6.3) that

$$\underset{\sim}{Y}_d^1 = \underset{\sim}{A}_d \underset{\sim}{\eta} - \underset{\sim}{H}_d \underset{\sim}{\varepsilon}_d^2 + \underset{\sim}{\gamma}_d^1 \tag{6.4}$$

where the d x d matrix $\underset{\sim}{A}_d = [(\underset{\sim}{F}')^{-(d-1)}\underset{\sim}{H}(1), \ldots, (\underset{\sim}{F}')^{-1}\underset{\sim}{H}(d)]$ where $(\underset{\sim}{F}')^{-S}$ denotes the matrix $[(\underset{\sim}{F}')^{-1}]^S$, the d x (d−1)d matrix

$$\underset{\sim}{H}_d = \begin{bmatrix} \underset{\sim}{H}'(1)\underset{\sim}{F}^{-1} & \underset{\sim}{H}'(1)\underset{\sim}{F}^{-2} & \cdots & \underset{\sim}{H}'(1)\underset{\sim}{F}^{-(d-1)} \\ \underset{\sim}{0} & \underset{\sim}{H}'(2)\underset{\sim}{F}^{-1} & \cdots & \underset{\sim}{H}'(2)\underset{\sim}{F}^{-(d-2)} \\ \vdots & \vdots & & \vdots \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{H}'(d-1)\underset{\sim}{F}^{-1} \\ \underset{\sim}{0} & \underset{\sim}{0} & \cdots & \underset{\sim}{0} \end{bmatrix},$$

the vector $\underset{\sim}{\varepsilon}_d^2 = [\underset{\sim}{\varepsilon}'(2), \ldots, \underset{\sim}{\varepsilon}'(d)]'$, and $\underset{\sim}{\gamma}_d^1 = [\gamma(1), \ldots, \gamma(d)]'$. Then, assuming $\underset{\sim}{A}_d$ is nonsingular, the transformation approach estimate of $\underset{\sim}{\beta}(d)$ based upon $\underset{\sim}{Y}_d^1$ is

$$\hat{\underset{\sim}{\beta}}(d|d) = \underset{\sim}{A}_d^{-1}\,\underset{\sim}{Y}_d^1 \quad . \tag{6.5}$$

It is easily verified that

$$\text{Var}\,[\underset{\sim}{\beta}(d) - \hat{\underset{\sim}{\beta}}(d|d)] = \underset{\sim}{A}_d^{-1}(\underset{\sim}{\Omega}_d + \sigma_\gamma^2\,\underset{\sim}{I}_d)(\underset{\sim}{A}_d^{-1})' \tag{6.6}$$

where the (i,j)th element of $\underset{\sim}{\Omega}_d = \text{Var}(\underset{\sim}{H}_d\underset{\sim}{\varepsilon}_d^2\underset{\sim}{H}_d')$ is given by

$$\Omega_{ij} = \begin{cases} \sum_{\ell=1}^{d-i} \underset{\sim}{H}'(i)\underset{\sim}{F}^{-\ell}\Sigma_\varepsilon(\underset{\sim}{F}')^{-(\ell+i-j)}\underset{\sim}{H}(j) & \text{if } 1 \le j \le i \le d-1 \\ 0 & \text{if } 1 \le j \le i = d \\ \Omega_{ji} & \text{if } 1 \le i \le j \le d \end{cases}$$

The initialization (6.5) and (6.6) is easily computed, and the Kalman filter can then be used with the remaining data to compute transformation approach estimates of the state vector $\underset{\sim}{\beta}(t)$ for likelihood evaluation when estimating σ_γ and $\underset{\sim}{\Sigma}_\varepsilon$, or for prediction.

7. Extension to the Case of Missing Data in the First d Time Points

If Y(t) is missing for one or more time points in t = 1, ..., d, then we do not have all the rows of the $\underset{\sim}{A}_d$ matrix defined in Section 3, and we cannot simply use data observed up to time d to eliminate $\underset{\sim}{\eta}$. We can use the first d observed data points, $Y(t_1)$, ..., $Y(t_d)$, to eliminate the effects of $\underset{\sim}{\eta}$ if $\underset{\sim}{A}_d$ =

$[\underset{\sim}{A}(t_1), \ldots, \underset{\sim}{A}(t_d)]'$ is nonsingular. But this $\underset{\sim}{A}_d$ may be singular depending on the pattern of missing data and the model in section 5, or depending on the independent variables and the model in section 6.

To handle the general case of $\underset{\sim}{A}_d$ singular, we need to find the first m for which $\underset{\sim}{A}_m = [\underset{\sim}{A}(t_1), \ldots, \underset{\sim}{A}(t_m)]'$ has full column rank. Then collecting the relations (2.4) for $t = t_1, \ldots, t_m$ (assume $t_m \le n$) we have

$$\underset{\sim}{Y}_m^1 = \underset{\sim}{A}_m \underset{\sim}{\eta} + \underset{\sim}{\omega}_m^1 \tag{7.1}$$

where $\underset{\sim}{A}_m$ is m x d and has rank $d \le m$. We apply the transformation approach directly to $\underset{\sim}{Y}_m$.

To do this let $\underset{\sim}{A}_m = \underset{\sim}{Q} \begin{bmatrix} \underset{\sim}{R} \\ \underset{\sim}{0} \end{bmatrix}$ be the Q-R decomposition of $\underset{\sim}{A}$ (Dongarra, et. al 1979, Chapter 9) where $\underset{\sim}{Q}$ is an m x m orthogonal matrix and $\underset{\sim}{R}$ is a d x d nonsingular upper triangular matrix. Partition $\underset{\sim}{Q}$ as $[\underset{\sim}{Q}_1 \; \underset{\sim}{Q}_2]$ where $\underset{\sim}{Q}_1$ is m x d and $\underset{\sim}{Q}_2$ is m x (m-d). We define a transformation from $\underset{\sim}{Y}_m^1$ to $\underset{\sim}{Z}_1$ and $\underset{\sim}{Z}_2$:

$$\begin{bmatrix} \underset{\sim}{Z}_1 \\ \underset{\sim}{Z}_2 \end{bmatrix} = \begin{bmatrix} (\underset{\sim}{A}_m' \underset{\sim}{A}_m)^{-1} \underset{\sim}{A}_m' \\ \underset{\sim}{Q}_2' \end{bmatrix} \underset{\sim}{Y}_m^1 = \begin{bmatrix} \underset{\sim}{R}^{-1} \underset{\sim}{Q}_1' \\ \underset{\sim}{Q}_2' \end{bmatrix} \underset{\sim}{Y}_m^1 \tag{7.2}$$

Here $\underset{\sim}{Z}_1 = \underset{\sim}{\eta} + (\underset{\sim}{A}_m'\underset{\sim}{A}_m)^{-1}\underset{\sim}{A}_m'\underset{\sim}{\omega}_m^1$, and $\underset{\sim}{Z}_2 = \underset{\sim}{Q}_2'\underset{\sim}{W}_m^1$ does not depend on $\underset{\sim}{\eta}$. (Note $\underset{\sim}{Q}_2'\underset{\sim}{A}_m = \underset{\sim}{Q}_2'\underset{\sim}{Q}_1\underset{\sim}{R} = \underset{\sim}{0}$.) From (2.3) we can initialize the Kalman filter at $t=t_m$ with

$$\hat{\underset{\sim}{X}}(t_m|t_m) = \underset{\sim}{\Phi}(t_m)(\underset{\sim}{A}_m'\underset{\sim}{A}_m)^{-1} \underset{\sim}{A}_m' \underset{\sim}{Y}_m^1 + E\,[\underset{\sim}{\xi}(t_m)|\underset{\sim}{Z}_2], \tag{7.3}$$

$$\mathrm{Var}(\underset{\sim}{X}(t_m) - \hat{\underset{\sim}{X}}(t_m|t_m)) = \mathrm{Var}(\underset{\sim}{\xi}(t_m)|\underset{\sim}{Z}_2) \tag{7.4}$$

where $\underset{\sim}{\xi}(t_m) = \underset{\sim}{\nu}(t) - \underset{\sim}{\Phi}(t)(\underset{\sim}{A}_m' \underset{\sim}{A}_m)^{-1}\underset{\sim}{A}_m' \underset{\sim}{\omega}_m^1$. The analogues to Results 2 and 3 follow immediately (with $\underset{\sim}{Z}_1$ replacing $\underset{\sim}{Y}_d^1$ in Assumption A). The QR decomposition simplifies the computations since $(\underset{\sim}{A}_m'\underset{\sim}{A}_m)^{-1}\underset{\sim}{A}_m' = \underset{\sim}{R}^{-1}\underset{\sim}{Q}_1'$.

The Kalman filter initialized with (7.3) and (7.4) produces transformation approach estimates $\hat{X}(t|t)$, $\hat{X}(t+1|t)$, and their error variances, only for $t \geq t_m$. This is sufficient for problems such as prediction or signal extraction that use all the data Y_n^1. If $m > d$, use of (7.3) and (7.4) leaves out some terms needed for likelihood evaluation as in Kohn and Ansley (1986). This can be accounted for by augmenting the transformation (7.2) to include

$$Z_3 = Y_n^{m+1} - \begin{bmatrix} A'(t_{m+1}) \\ \vdots \\ A'(t_n) \end{bmatrix} (A_m' A_m)^{-1} A_m' Y_m^1$$

The desired likelihood function is $p(Z_2, Z_3) = p(Z_2)p(Z_3|Z_2)$ where $p(Z_2, Z_3)$, $p(Z_2)$ and $p(Z_3|Z_2)$ are respectively the joint density of Z_2 and Z_3, the marginal density of Z_2, and the conditional density of Z_3 given Z_2. The Kalman filter initialized with (7.3) and (7.4) can be used to compute $p(Z_3|Z_2)$, so we must additionally compute $p(Z_2)$ directly, which is not computationally burdensome unless m is large relative to d. Notice that if m=d then Z_2 disappears from (7.2) and use of the Kalman filter initialized by (7.3) and (7.4) is all we need.

It is difficult to give general expressions for the evaluation of (7.3) and (7.4) since the form of the results will depend on the model involved and the location of the missing data, etc. Specific problems are best handled on a case by a case basis. The ability to directly handle arbitrary patterns of missing data and the case of A_d singular is the primary advantage of the modified Kalman filter of AK.

Note on Transformation Approach Likelihoods

AK (Theorem 5.1) makes an important point about transformation approach likelihoods, namely, that they are defined only up to multiplicative constants. For this reason, we did not bother to include the Jacobian of the transformation (7.2), which is $|\underset{\sim}{R}|^{-1}$, in the likelihood defined above. The same consideration also applies to the previous case of $\underset{\sim}{A}_d$ nonsingular, for which the Jacobian of the transformation given in section 3 would be $|\underset{\sim}{A}_d|^{-1}$. Excluding $|\underset{\sim}{A}_d|^{-1}$ from the likelihood for the ARIMA component models of section 5, and using the initialization of Result 4, results in precisely the same likelihood as that for the differenced data. This points out that the likelihoods are invariant to different transformations or, conversely, to different choices of starting values to eliminate, only up to the multiplicative constants. As long as $\underset{\sim}{A}_m$ does not depend on model parameters (as assumed by AK), leaving out the Jacobian poses no problem for model estimation. If $\underset{\sim}{A}_m$ does depend on model parameters, however, this matter may merit more thought. (Such dependence occurs in ARIMA component models with AR operators handled as discussed at the end of section 5 to avoid assuming stationarity.) Even so, at least for the case of $\underset{\sim}{A}_d$ nonsingular, we can avoid problems by choosing to always eliminate starting values $\underset{\sim}{Y}_d{}^1$, since the Jacobian of the transformation analogous to that given in section 3 will then not depend on the model parameters. The results are again the same as using the initialization of result 4 and excluding $|\underset{\sim}{A}_d|^{-1}$ from the likelihood.

8. Example

Here we present an example to illustrate the potential differences between the initialization in this paper and more naive initializations as are discussed in the introduction. In Bell and Hillmer (1984, pp. 310-313) we considered the model $(1-.26B)(1-B)(1-B^{12})Y(t) = (1-.88B^{12})a(t)$ with $\sigma_b^2 = 16150$ for the time series of employed males aged 20 and older in nonagricultural industries, using data from the U.S. Bureau of Labor Statistics for the period January 1965 through August 1979(176 observations). The "canonical decomposition" considered there has Y(t) = S(t) + N(t), with the seasonal and nonseasonal components following the models:

$$(1 + B + \cdots + B^{11})S(t) = \theta_s(B)b(t) \qquad \sigma_b^2 = 82.11$$

$$(1 - .26B)(1 - B)^2 N(t) = (1 - .990B + .001B^2)c(t) \qquad \sigma_c^2 = 14412$$

where $\theta_s(B)$, a polynomial in B of degree 11, is given in the paper. These component models are consistent with the model given for Y(t). Subsequently, Burridge and Wallis (1985) considered this example to illustrate how the Kalman filter/smoother could be used to produce signal extraction variances for seasonal adjustment. They initialized the Kalman filter at the first time point (t=1 in our notation, t=0 in theirs) with an initial state estimate of $\underset{\sim}{0}$ and a "large" initial variance of $P(1|0) = 10^{12}I$.

Figure 1 shows how the initialization affects the innovation variances, Var(Y(t)|t-1), from the Kalman filter. Using the Kalman filter with the component models and the initialization (5.13)-(5.14) produces the same results as using the ARIMA model for Y(t) and the initialization (5.15)-(5.16), or the modified Kalman filter of AK suitably initialized, or using the stationary model for the differenced data $(1-B)(1-B^{12})Y(t)$ and the usual initialization of the Kalman filter in the stationary case. The resulting $\log_{10}\mathrm{Var}(Y(t)|t-1)$ for t = 14, ..., 176 are shown as the dotted line in Figure 1. The corresponding results with the initialization used by Burridge and Wallis are the solid line in Figure 1. There appear to be important differences in the results, especially in the first half of the series. Interestingly, if the initialization $P(0|0) = 10^{12}I$ is used instead and compared to the results from the initialization of this paper, visually a pattern similar to that in Figure 1 appears, but the magnitude of the difference appears only about half as great. This shows that the choice between alternative ways of initializing with a "large" covariance matrix is not as innocuous as it may seem.

REFERENCES

Akaike, H. (1978) "Covariance matrix computation of the state variable of a stationary Gaussian process," Annals of the Institute of Statistical Mathematics, Vol. 30, Part B, 499-504.

Ansley, C. F. and Kohn, R. (1985a) "Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions," Annals of Statistics, 13, p.1286-1316.

_______(1985b) "A Structured State Space Approach to Computing the Likelihood of an ARIMA Process and Its Derivatives," Journal of Statistical Computation and Simulation, 21, 135-169.

Bell, W. R. (1982) "Signal extraction for nonstationary time series," Report Number 82/01, Statistical Research Division U.S. Bureau of the Census.

________(1984) "Signal extraction for nonstationary time series," Annals of Statistics, 13, p. 646-664.

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

Burridge, P. and Wallis, K. F. (1985) "Calculating the variance of seasonally adjusted series," Journal of the American Statistical Association, 80, p.541-552.

De Jong (1988) "The likelihood for a state space model," Biometrika, 75, p. 165-169.

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.

Harrison, P. J. and Stevens, C. F. (1976) "Bayesian forecasting," Journal of the Royal Statistical Society B, 38, p. 205-247.

Harvey, A. C. (1981) "Finite sample prediction and overdifferencing" Journal of Time Series Analysis, 2, p. 221-232.

________ and Phillips (1979), "Maximum likelihood estimation of regression models with autoregressive-moving average disturbances," Biometrika, 66, p. 49-58.

Harvey, A. C. and Pierse, R. G. (1984), "Estimating missing observations in economic time series," Journal of the American Statistical Association, 79,

Jones, R. H. (1980) "Maximum likelihood fitting of ARMA models to time series with missing observations," Technometrics, 22, p.389-395.

________ (1985) "Time series analysis with unequally spaced data," Handbook of Statistics Vol 5, ed. E J. Hannan, P. R. Krishnaiah and M. M. Rao, Elsevier Science, p.157-177.

Kitagawa, G. (1981) "A nonstationary time series model and its fitting by a recursive filter," Journal of Time Series Analysis, 2, p.103-116.

Kohn, R. and Ansley C. F. (1986) "Estimation, prediction, and interpolation for ARIMA models with missing data," Journal of the American Statistical Association, 81, p.751-761.

Kohn, R. and Ansley, C. F. (1987) "Signal extraction for finite nonstationary time series," Biometrika, 74, p.411-421.

Machak, J.A., Spivey, W.A., and Wrobleski, W.J. (1985) "A Framework for Time Varying Parameter Regression Modeling," Journal of Business and Economic Statistics, 3, 104-111.

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

Var(Y(t|t-1)), t=14,...,176 for Model 2 of Burridge and Wallis
Solid- Naive Initialization: P(1|0) = 10**12 I
Dotted- Transformation Approach Estimate of P(13|13)