BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-85/10

AN ANALYSIS AND GENERALIZATION OF M. WATSON'S MINIMAX PROCEDURE FOR DETERMING THE COMPONENT ESTIMATES OF A SEASONAL TIME SERIES

by

David F. Findley
Statistical Research Division
Bureau of the Census
Washington, D.C. 20233

Recommended by: Nash J. Monsour

Report completed: June 27, 1985 (REVISED)

Report issued: July 29, 1985

Revision No. 1 : June 27 , 1985

An Analysis and Generalization of M. Watson's Minimax Procedure for Determining the Component Estimates of a Seasonal Time Series

David F. Findley
Statistical Research Division
Census Bureau
Washington, D.C. 20233

AUTHOR'S FOOTNOTE

David F. Findley is a Principal Researcher in the Statistical Research Division of the Census Bureau, Washington, DC 20233. Part of this work was completed during the author's stay at the University of Heidelberg sponsored by the Sonderforschungsbereich 123 of the Deutsche Forschungsgemeinschaft. The author wishes to express his gratitude for this support and for the helpful comments on drafts of this paper from William Bell, J.P. Burman, Jürgen Franke, Mark Watson and the referees. The remarks of one referee were particularly stimulating. We also thank Brooks Ferebee for computational assistance.

ABSTRACT

It is shown that the solution trichotomy described by Watson (1984) applies to a much larger class of minimax problems and has additional robustness properties beyond those discussed by him. We also address a number of practical issues, demonstrating, in particular, that the fixed–interval smoothing algorithm can be used, when finite and possibly nonconsecutive data are available, to obtain the minimax solution, the corresponding component estimate, and the robust mean square estimation error.

KEYWORDS. Signal extraction; Nonstationary time series; Seasonal adjustment; Robust methods; Missing data.

1. INTRODUCTION

In the Hillmer–Tiao minimum mean square signal extraction approach to seasonal adjustment, see Hillmer, Bell and Tiao (1983) or Burman (1980), the observed series is modeled as a nonstationary seasonal ARMA process x_t having a decomposition of the form

$$x_t = n_t + s_t + e_t \,, \tag{1.1}$$

in which s_t and n_t are "minimum variance" seasonal and nonseasonal components, and e_t is a white noise process (usually nonzero) admitting a continuous family of decompositions

$$e_t = e_t^\alpha + \bar{e}_t^\alpha \,,\ m \leqq \alpha \leqq M \,, \tag{1.2}$$

into uncorrelated white noise series e_t^α and $\bar{e}_t^\alpha$ with variances $\alpha - m$ and $M - \alpha$, respectively. Thus, by means of the definitions

$$\begin{aligned} n_t^\alpha &= n_t + e_t^\alpha \\ s_t^\alpha &= s_t + \bar{e}_t^\alpha \,, \end{aligned} \tag{1.3}$$

a family of possible nonseasonal–seasonal component decompositions of x_t is obtained,

$$x_t = n_t^\alpha + s_t^\alpha \,,\ m \leqq \alpha \leqq M \,, \tag{1.4}$$

arising from possible apportionments of e_t.

Given some set of x–variables, say $\{x_{t-j},\ j \in J\}$, let $\hat{n}_t^\alpha = \hat{n}_t^\alpha(J)$ denote the linear function of the given variables for which the mean square error

$$E\{n_t^\alpha - \hat{n}_t^\alpha\}^2 \tag{1.5}$$

is minimized. [1] The determination of such a component estimate is what is meant by signal extraction, but complementary to the task of determining an optimal estimate of the nonseasonal component of x_t, there is the problem of deciding on a choice of α. The authors referred to above prefer α=M, arguing that the decomposition of x_t with the most predictable seasonal component should be used (which sometimes does not force the choice α=M, see section 6 below.) On the other hand, many data users prefer to have the nonseasonal component be as predictable as possible, which suggests the choice α=m. Watson (1984) has proposed two minimax criteria for making this choice, leading to an estimate $\hat{n}_t^{\beta^*}$, where $\beta^*=\beta^*(t)$ solves

$$\beta^* : \min_{m\leq\beta\leq M} \max_{m\leq\alpha\leq M} E\{n_t^\alpha - \hat{n}_t^\beta\}^2 , \tag{1.6}$$

or to $\hat{n}_t^{\beta^{**}}$, where $\beta^{**}=\beta^{**}(t)$ solves the analogue of (1.6) for first differences,

$$\beta^{**} : \min_{m\leq\beta\leq M} \max_{m\leq\alpha\leq M} E\{\Delta n_t^\alpha - \Delta\hat{n}_t^\beta\}^2 . \tag{1.7}$$

Watson gives an analysis of (1.6) and (1.7) only for the situation in which complete bi-infinite data are available ($J = (-\infty,\infty) =_{\mathrm{def}} \{0,\pm1, ...\}$) and the transfer function of the filter determining the estimates $\hat{n}_t^\beta$ is given by the pseudo-spectrum ratio analogue of the classical Wiener-Kolmogoroff formula (see Appendix 1), as under Assumption A of Bell (1984). He demonstrates for (1.6), and analogously for (1.7), that one of three possible numbers, m, M and β_0, in our later notation, is the solution (actually he doesn't discuss uniqueness), observing that $E\{n_t^{\beta_0} - \hat{n}_t^{\beta_0}\}^2$, when applicable, provides a robust estimate of the mean squared estimation error.

Watson's very stimulating paper leaves a number of practical questions unanswered concerning the solution of (1.6) (and (1.7)) : (i) Does the same kind of m, M, β_0 trichotomy occur when there are only finitely many, possibly nonconsecutive observations ? (ii) Is the solution always a function just of the coefficients of the optimal filter ? (iii) If a state space smoothing algorithm is used for the determination of $\hat{n}_t^\beta$, so that filter coefficients are not directly available, can the solution be determined from quantities calculated by the algorithm ? (iv) Is there a sense in which the quantity $E\{n_t^{\beta^*} - \hat{n}_t^{\beta^*}\}^2$ is always a robust estimate of mean square estimation error ? (v) How can $E\{n_t^{\beta^*} - \hat{n}_t^{\beta^*}\}^2$ be calculated ? (vi) Do the same results hold if the e_t^α, m$\leq\alpha\leq$M are, say, moving average processes ? In work in progress at the Census Bureau, J.P. Burman has begun exploring the use of MA(1) components e_t in (1.1), see Appendix 1 for some considerations motivating this possibility.

The goal of this paper is to provide answers to these questions and others. In outline, its content is as follows. An elementary minimax problem is discussed in section 2 and shown, in sections 3 and 4, to cover very general cases of (1.6) and (1.7) and to provide an affirmative answer to questions (i), (iv) and (vi). Questions (iii) and (v) are addressed in subsection 3.2, and (ii) is treated in subsection 3.3. In section 5, it is shown that our general assumptions (4.2) and (3.3) are satisfied for the most familiar category of nonstationary time series under Assumption A of Bell (1984). A simple but quite illuminating example is analyzed in section 6.

A comment concerning the significance of (i) is in order. It could turn out that Watson's solution of (1.6) (or (1.7)) for the case of bi-infinite data provides an adequate approximation to the exact solution for the finite data situation, for most of the observation times, for many series. Of course, such an empirical issue cannot be investigated without a method which provides the exact answer (under the same model assumptions). Nevertheless, it would be surprising if the bi-infinite solution consistently provided good approximations near the ends of the series. This, however, is precisely where recent observations lie, which are usually the observations of most interest to consumers of seasonally adjusted data. It seems important, therefore, to have a solution method for the case of finitely many observations.

2. THE SOLUTION OF A SPECIAL QUADRATIC MINIMAX PROBLEM

In the next two sections of this paper, we will be interested in determining the (unique) value β^* of β solving the minimax problem

$$\beta^* : \min_{m \leq \beta \leq M} \max_{m \leq \alpha \leq M} F(\alpha,\beta) \tag{2.1}$$

for given real numbers $m<M$ and for what will be shown to be special cases of the linear–quadratic function

$$F(\alpha,\beta) = A+D(\beta-m)^2+(\alpha-m)(C-2D(\beta-m)) \quad (D>0) \; . \tag{2.2}$$

The value β_0 of β for which the linear function $C(\beta) = C-2D(\beta-m)$ is zero will turn out to be of particular interest. Clearly

$$\beta_0 = m + C/2D\ , \tag{2.3}$$

or, if $C(\alpha)$ and $C(\beta)$ are known for two values, α and β ,

$$\beta_0 = \beta - (\beta-\alpha)\ C(\beta)\ (C(\beta)-C(\alpha))^{-1}\ . \tag{2.4}$$

The assertions (2.5) – (2.7) below are verified in Appendix 2.

(2.5) *For each fixed* α, *the unique value of* β *minimizing* $F(\alpha,\beta)$ *over any interval containing* α *is* $\beta = \alpha$.

(2.6) *There is a unique solution* β^* of (2.1). *It is that number in the interval* $[m,M]$ *which is closest to the number* β_0 *defined by* (2.3). *That is,* $\beta^* = \beta_0$, m *or* M, *depending on whether* (i) $m \leqq \beta_0 \leqq M$, (ii) $\beta_0 < m$ *or* (iii) $\beta_0 > M$ *holds. Since* $C(\beta)$ *is a decreasing function of* β, *these conditions are equivalent to* (i)' $C(M) \leqq 0 \leqq C(m)$, (ii)' $C(m)<0$ *and* (iii)' $C(M)>0$, *respectively. In case* (i), *therefore,* $F(\alpha,\beta^*)$ *is independent of* α .

(2.7) *The solution* β^* *of* (2.1) *described in* (2.6) *is also the unique value of* α *maximizing* $F(\alpha,\alpha)$ *over* $[m,M]$ *and we have*

$$\max_{m\leq\alpha\leq M} F(\alpha,\alpha) = F(\beta^*,\beta^*) = \min_{m\leq\beta\leq M} \max_{m\leq\alpha\leq M} F(\alpha,\beta)\ . \tag{2.8}$$

3. ANALYSIS OF (1.6) IN A GENERAL CONTEXT

We shall now show that the mean square expression in (1.6) always has the form (2.2) when certain covariances associated with the random variable e_t^α used to define $n_t^\alpha = n_t + e_t^\alpha$ are proportional to $(\alpha-m)$. By (2.6) – (2.8), it follows that (1.6) has a unique solution $\beta^*(t)$ which defines a unique robust estimate $E\{n_t^{\beta^*(t)} - \hat{n}_t^{\beta^*(t)}\}^2$ of the mean square estimation error. We also describe a practical procedure for calculating this robust estimate and methods for calculating $\hat{n}_t^{\beta^*(t)}$ when ARMA models are available for x_t, n_t, e_t^α and $s_t^\alpha = x_t - e_t^\alpha$ for some $\alpha>m$. However, the initial discussion of (1.6) relies only on general properties of least mean square approximations and has no necessary connection with time series or time series models. (The subscript t is fixed and could be omitted.)

Our assumptions are that for some finite or infinite set of integers J, the random variables n_t, e_t^α, $m \leqq \alpha \leqq M$ and x_{t-j}, $j \in J$ are defined on the same probability space, have mean 0 and finite variance(s), and have the following three additional properties :

$$\mathbf{var}(e_t^\alpha) = (\alpha - m)d(t,t)\ , \quad m \leqq \alpha \leqq M \tag{3.1}$$

for some $d(t,t) > 0$,

$$\mathbf{cov}(n_t, e_t^\alpha) = (\alpha - m)f(t,t)\ , \quad m \leqq \alpha \leqq M \tag{3.2}$$

for some $f(t,t)$, and

$$\mathbf{cov}(x_{t-j}, e_t^\alpha) = (\alpha - m)g(t-j,t)\ , \quad m \leqq \alpha \leqq M \tag{3.3}$$

for some $g(t-j,t)$ defined for all $j \in J$.

In the nonstationary ARMA model signal extraction context, we shall show in section 5 that special cases of (3.2) and (3.3) are inevitable consequences of (3.1) and Assumption A of Bell (1984). It will also be observed in section 6 that the traditional assumption of uncorrelated components, suggested by the pseudospectrum decompositions described in Appendix 1, which also implies special cases of (3.2) and (3.3), is less convenient than Assumption A.

3.1 Solution of (1.6)

Let OBS = OBS(J) denote the space of all linear combinations of x_{t-j}, $J \in J$, together with their mean square limits if J is infinite. For any random variable y with mean zero and finite variance of the sort to be considered below, we will denote by $\hat{y}$ the least mean square approximation to y in OBS, and recall that it is characterized by the property that $y - \hat{y}$ is uncorrelated with every random variable in OBS. Therefore, if we have two such random variables y and z, then $\hat{y} + \hat{z}$ is the least mean square approximation in OBS to y+z. Also, note that if $\mathbf{cov}(y, x_{t-j}) = \mathbf{cov}(z, x_{t-j})$, $j \in J$, then $y - z$ is uncorrelated with OBS, so that $\hat{y} - \hat{z} = 0$, i.e., $\hat{y}=\hat{z}$. By using this argument, it follows from (3.3) that if we set $\tilde{e}_t = (M-m)^{-1}\hat{e}_t^M$, then

$$\hat{e}_t^\alpha = (\alpha-m)\tilde{e}_t\ , \quad m\leqq\alpha\leqq M\ . \tag{3.4}$$

Thus (3.3) has the important consequence that $\hat{n}_t^\alpha = \hat{n}_t + \hat{e}_t^\alpha$ is a *linear* function of α,

$$\begin{aligned} \hat{n}_t^\alpha &= \hat{n}_t + (\alpha-m)\tilde{e}_t \\ &= \hat{n}_t^\beta + (\alpha-\beta)(\beta-\gamma)^{-1}\{n_t^\beta-n_t^\gamma\}\ , \quad m\leqq\alpha\leqq M\ , \end{aligned} \tag{3.5}$$

for any distinct pair β,γ in [m,M]. We shall also assume

$$E\tilde{e}_t^2 \neq 0\ . \tag{3.6}$$

Otherwise we would have $\hat{n}_t^\alpha = \hat{n}_t$ for all α, and (1.6) would be trivial.

Since $\mathrm{cov}(n_t-\hat{n}_t,\ e_t^\alpha-\hat{e}_t^\alpha) = \mathrm{cov}(n_t,e_t^\alpha) - \mathrm{cov}(\hat{n}_t,\hat{e}_t^\alpha)$, the formulas (3.2) and (3.4) imply that

$$\begin{aligned} \mathrm{cov}(n_t-\hat{n}_t,\ e_t^\alpha-\hat{e}_t^\alpha) &= (\alpha-m)\ \{f(t,t) - \mathrm{cov}(\hat{n}_t,\tilde{e}_t)\} \\ &= (\alpha-m)(\beta-m)^{-1}\ \mathrm{cov}(n_t-\hat{n}_t,\ e_t^\beta-\ \hat{e}_t^\beta)\ . \end{aligned} \tag{3.7}$$

Also, since $E\{e_t^\alpha-\hat{e}_t^\alpha\}^2 = E\{e_t^\alpha\}^2 - E\{\hat{e}_t^\alpha\}^2$, we obtain from (3.1) and (3.4) that

$$E\{e_t^\alpha - \hat{e}_t^\alpha\}^2 = (\alpha-m)d(t,t) - (\alpha-m)^2\ E\tilde{e}_t^2\ . \tag{3.8}$$

These are the key formulas for the analysis of $E\{n_t^\alpha-\ \hat{n}_t^\beta\}^2$, $m\leqq\alpha,\beta\leqq M$. We can now show that

$$E\{n_t^{\alpha}-\hat{n}_t^{\beta}\}^2 = E\{n_t-\hat{n}_t\}^2 + (\beta-m)^2E\tilde{e}_t^2 + (\alpha-m)\{C_t-2E\tilde{e}_t^2\,(\beta-m)\} \tag{3.9}$$

with

$$C_t = d(t,t) + 2\{f(t,t) - \mathrm{cov}(\hat{n}_t, \tilde{e}_t)\}\ . \tag{3.10}$$

Indeed, from the decomposition

$$n_t^{\alpha}-\hat{n}_t^{\beta} = \{n_t-\hat{n}_t\} + \{e_t^{\alpha}-\hat{e}_t^{\alpha}\} + (\alpha-\beta)\tilde{e}_t$$

we obtain

$$E\{n_t^{\alpha}-\hat{n}_t^{\beta}\}^2 = E\{n_t-\hat{n}_t\}^2+E\{e_t^{\alpha}-\hat{e}_t^{\alpha}\}^2+(\alpha-\beta)^2E\tilde{e}_t^2 + 2\ \mathrm{cov}(n_t-\hat{n}_t,\ e_t^{\alpha}-\hat{e}_t^{\alpha})$$

and (3.9) and (3.10) follow readily from (3.7) and (3.8).

The formulas (3.9) and (3.10) show that the minimax problem (1.6) is a special case of the problem (2.1) and (2.2) already solved in section 2. Of course, we must describe how, in practice, one can obtain the following three quantities : (i) the pivotal value

$$\beta_0(t) = m + C_t/\{2E\tilde{e}_t^2\}\ , \tag{3.11}$$

which identifies the unique solution $\beta^*(t)$ of (1.6) as in (2.6); (ii) the minimax robust mean square error measure

$$E\{n_t^{\beta^*(t)}-\hat{n}_t^{\beta^*(t)}\}^2 = E\{n_t-\hat{n}_t\}^2+E\{e_t^{\beta^*(t)}-\hat{e}_t^{\beta^*(t)}\}^2 + 2\mathrm{cov}(n_t-\hat{n}_t, e_t^{\beta^*(t)}-\hat{e}_t^{\beta(t)}) \tag{3.12}$$

and, (iii) whichever estimate $\hat{n}_t^{\alpha(t)}$ is desired. (Some analysts might prefer a choice $\hat{n}_t^{\alpha(t)}$ with $\alpha(t) \neq \beta^*(t)$ but still wish to use the robust measure (3.12)). There is a practical and quite general context in which all of these quantities are readily calculated, which we describe in the next subsection.

3.2 Utilizing State Space Smoothing Algorithms

Suppose that observations $x_{t_1}, \ldots, x_{t_N}$ ($t_1<t_2< \ldots <t_N$) are available, along with a state space representation of the process x_t, $t\geqq t_1$, which satisfies the conditions permitting the use of the Kalman Filter, and whose state vector contains, say, $n_t = n_t^m$ and either e_t^{β} or $s_t^{\beta} = x_t-n_t-e_t^{\beta}$ for some $\beta\neq m$ among its entries (see also

subsection 3.4). This can usually be accomplished in a variety of ways when ARMA models (stationary or not) are available for the series x_t, n_t, e_t^β and s_t^β, as several of the papers in Parzen (1984) demonstrate.

Then the minimum mean square error estimates $\hat{n}_t$, $\hat{e}_t^\beta$ and $\hat{s}_t^\beta$ in the space OBJ spanned by $\{x_{t_1}, \ldots, x_{t_N}\}$ can be calculated for $t \in \{t_1, \ldots, t_N\}$ by the familiar state space smoothing algorithms (Anderson and Moore, 1979; Parzen 1984). These algorithms also produce the associated estimation error variances and covariances. Even when e_t^β is not a state vector entry (as happens in the example of section 6 below), we can use the fact that $t \in \{t_1, \ldots, t_N\}$ implies

$$n_t + e_t^\beta + s_t^\beta = x_t = \hat{n}_t + \hat{e}_t^\beta + \hat{s}_t^\beta$$

to calculate $\hat{e}_t^\beta = x_t - \hat{n}_t - \hat{s}_t^\beta$ and

$$e_t^\beta - \hat{e}_t^\beta = -(n_t - \hat{n}_t) - (s_t^\beta - \hat{s}_t^\beta) . \tag{3.13}$$

From $\hat{n}_t$ and $\hat{e}_t^\beta$, we can obtain $\hat{n}_t^\alpha$ for all α using (3.5). Utilizing $E\{n_t - \hat{n}_t\}^2$, $E\{s_t^\beta - \hat{e}_t^\beta\}^2$ and $\mathrm{cov}(n_t - \hat{n}_t, s_t^\beta - \hat{s}_t^\beta)$, we can determine $E\{e_t^\beta - \hat{s}_t^\beta\}^2$ and $\mathrm{cov}(n_t - \hat{n}_t, e_t^\beta - \hat{e}_t^\beta)$ via (3.13) .

Having these quantities, we can calculate $E\tilde{e}_t^2$ from (3.8) with $\alpha=\beta$ and C_t from (3.7) and (3.9), and therefore $\beta_0(t)$ in (3.11). By (2.6), this determines the solution $\beta^*(t)$ of the minimax problem (1.6). Finally, if the robust measure $E\{n_t^{\beta^*(t)} - \hat{n}_t^{\beta^*(t)}\}^2$ has not already been obtained, that is, if $\beta^*(t) \neq m_t\beta$, then it can be derived from (3.12), using (3.7) and (3.8) with $\alpha=\beta^*(t)$ to evaluate the terms on the right hand side.

As the example of section 6 below shows, the development of the kind of state space representation needed for the anlysis described in this section requires a spectral factorization as well as some straightforward computations to determine the initializing covariance matrix. The payoff, as we have just seen, is that then a single run of the fixed-interval smoothing algorithm over $t_1 \leqq t \leqq t_N$ yields the solution to the minimax problem, and also yields the associated estimates and robust mean square error measures, for all $t \in \{t_1, \ldots, t_N\}$.

3.3 Utilizing Filter Coefficients

Another approach utilizes the coefficients $h^{\alpha}_{j,t}$, $h^{\beta}_{j,t}$, $j \in J$ $(\alpha \neq \beta)$ of filters determining two estimates $\hat{n}^{\alpha}_t$ $(= \Sigma_{j\in J}\, h^{\alpha}_{j,t}\, x_{t-j})$ and $\hat{n}^{\beta}_t$ as functions of x_{t-j}, $j \in J$, for each t of interest. Given such information, it follows from (3.5) that all estimates and their associated filters can be readily obtained. We shall show in this section that the filter coefficient information can be combined with the covariance information (3.1) – (3.3) to obtain the pivotal value $\beta_0(t)$, and hence the solution $\beta^*(t)$, very simply in the most typical situations. This approach does not immediately produce the minimax robust mean square error measure $E\{n^{\beta^*(t)}_t - \hat{n}^{\beta^*(t)}_t\}^2$, however, and it also has the practical disadvantage that there do not appear to be any attractive computational algorithms known at present for obtaining these coefficients for finite observation sets $\{x_{t-j} : j \in J\} = \{x_{t_i} : 1 \leq i \leq N\}$.

In analogy with the notation of section 2, let us denote the multiplier of $\alpha - m$ on the right in (3.9) by $C_t(\beta)$. Since $E\tilde{e}^2_t(\beta - m) = \mathrm{cov}(\hat{e}^{\beta}_t, \tilde{e}_t)$, by (3.4), it follows from (3.9) and (3.10) that

$$C_t(\beta) = d(t,t) + 2f(t,t) - 2\mathrm{cov}(\hat{n}^{\beta}_t, \tilde{e}_t) \,. \tag{3.14}$$

Since $\hat{n}^{\beta}_t = \Sigma_{j\in J}\, h^{\beta}_{t,j}\, x_{t-j}$ (mean square convergence if J is infinite), we have, from (3.3) and (3.5),

$$\mathrm{cov}(\hat{n}^{\beta}_t, \tilde{e}_t) = \Sigma_{j\in J}\, h^{\beta}_{t,j}\, g(t-j,t) \,. \tag{3.15}$$

Thus $C_t(\beta)$ and $C_t(\alpha)$ can be determined from the given filter coefficients and the values of $d(t,t)$ and $f(t,t)$. By (2.4), the pivotal value $\beta_0(t)$ determining the minimax solution $\beta^*(t)$ can be calculated as

$$\beta_0(t) = \beta - (\beta - \alpha)\, C_t(\beta)(C_t(\beta) - C_t(\alpha))^{-1} \,.$$

The simplest case is the one in which e^M_t, $-\infty < t < \infty$ is a white noise process which is uncorrelated with the other components of x_t. Then $d(t,t) = 1$ $f(t,t) = 0$ and

$$g(t-j,t) = \begin{cases} 0 & \text{if } j \neq 0 \\ 1 & \text{if } j=0 \end{cases} .$$

From these values, it follows via (3.14) – (3.15) that

$$C_t(\beta) = 1 - 2\, h^{\beta}_{t,0} . \tag{3.16}$$

If (3.16) holds, then, by (2.6), the ordering of $h^{m}_{t,0}$ and $h^{M}_{t,0}$ relative to 1/2 determine the solutions of (1.6) as in the special situation considered in Watson (1984). However, we shall see in section 6 that even when J = $(-\infty,\infty)$, the solution described by Watson can be incorrect for the intializing values of the model equations, because of the inapplicability of (3.16) there.

3.4 Intervals different from [m,M]

In our motivating discussion in Appendix 1, the interval [m,M] was described as the largest interval over which the process e^{α}_t could be defined. This property has played no role in our analysis, however. More precisely, let us, in the spirit of section 1, make the additional assumption that for any $\overline{m}$ in [m,M] and any $\alpha \geq \overline{m}$, the processes $e^{\overline{m}}_t$ and $\overline{e}^{\alpha}_t = e^{\alpha}_t - e^{\overline{m}}_t$ are uncorrelated. Then, for any $\overline{M}$ in $(\overline{m},M]$, if we replace the quantities m,M,e^{α}_t and n_t in all previous formulas by $\overline{m},\overline{M},\overline{e}^{\alpha}_t$ and $\overline{n}_t = n_t + e^{\overline{m}}_t$, valid results are obtained. In particular, $\beta_0(t)$ in (3.11) can be determined by an analysis over either $[\overline{m},\overline{M}]$ or [m,M]. For the solution methodology described in subsection 3.2, an alternative is, therefore, to have $n^{\overline{m}}_t$ and either $\overline{e}^{\beta}_t$ or s^{β}_t in the state vector, for some $\overline{m}>m$ and $\beta \in [\overline{m},\overline{M}]$.

4. ANALYSIS OF (1.7) IN A GENERAL CONTEXT

The fact that the best approximation in OBS to the first difference Δn^{α}_t is given by $\Delta \hat{n}^{\alpha}_t = \hat{n}^{\alpha}_t - \hat{n}^{\alpha}_{t-1}$ makes it possible to use the results of section 3 to solve (1.7) at the same level of generality used for (1.6). If we consistently replace e^{α}_t and n_t and their best approximations, in the assumptions (3.1) – (3.3), (3.6) and in the remaining formulas of subsections 3.1, 3.3 and 3.4, by the corresponding differences Δe_t^α, Δn_t etc., then the resulting formulas and assertions are correct. For the approach of subsection 3.2, replacing the entries in the state equation by their first differences yields a state equation appropriate for estimation from data Δx_t, but this is not what is wanted. To use the given data x_t, it is probably simplest to develop a state equation whose entries contain n_t, n_{t-1}, e_t^α and e_{t-1}^α, or entries yielding these quantities by linear combination.

We now present generalizations of (3.1) and (3.2) which ensure that their analogues hold when e_t^α and n_t are replaced by first differences. Suppose, for simplicity, that e_t^α and n_t are defined for all t. It then clearly suffices to assume that

$$\mathrm{cov}(e_r^\alpha, e_t^\alpha) = (\alpha-m)d(r,t)\ ,\quad m\leqq\alpha\leqq M \tag{4.1}$$

and

$$\mathrm{cov}(n_r, e_t^\alpha) = (\alpha-m)f(r,t)\ ,\quad m\leqq\alpha\leqq M \tag{4.2}$$

hold for some d(r,t) and f(r,t), for all r and t. From (4.1) and (4.2) we obtain the following analogues of (3.1) and (3.2) :

$$E\{\Delta e_t^\alpha\}^2 = (\alpha-m)\{d(t,t)+d(t-1,t-1)-2d(t,t-1)\}\ , \tag{4.3}$$

$$\mathrm{cov}(\Delta n_t,\Delta e_t^\alpha) = (\alpha-m)\{f(t,t)+f(t-1,t-1)-f(t-1,t)-f(t,t-1)\}\ . \tag{4.4}$$

From (3.3), we have

$$\mathrm{cov}(x_{t-j}, \Delta e_t^\alpha) = (\alpha-m)\{g(t-j,t)-g(t-j,t-1)\}\ . \tag{4.5}$$

The analysis of (1.7) then proceeds by replacing the quantities d(t,t), f(t,t) and g(t−j,t) in the analysis of (1.6) by the corresponding quantity on the right in (4.3), (4.4) and (4.5), respectively. (We must assume $\Delta\tilde{e}_t\neq0$. Of course (1.7) is trivial if $\Delta\tilde{e}_t=0$.) For example, under the assumptions used to derive (3.16), we have $g(r,r) = d(r,r) = 1$ for all t, $g(r,s) = d(r,s) = 0$ for $r\neq s$, and $f(r,s)\equiv0$ for all $r,s \in \{t-1,t\}$. Therefore, since

$$\mathrm{cov}(\Delta\hat{n}^{\beta}_{t}, \Delta\tilde{e}_{t}) = \mathrm{cov}(\hat{n}^{\beta}_{t}, \tilde{e}_{t})+\mathrm{cov}(\hat{n}^{\beta}_{t-1}, \tilde{e}_{t-1})-\mathrm{cov}(\hat{n}^{\beta}_{t}, \tilde{e}_{t-1})-\mathrm{cov}(\hat{n}^{\beta}_{t-1}, \tilde{e}_{t})\ ,$$

the analogue of (3.16) for (1.7) is

$$C^{\Delta}_{t}(\beta) = 2\{1-h^{\beta}_{t,0}-h^{\beta}_{t-1,0}+h^{\beta}_{t,1}+h^{\beta}_{t-1,-1}\}\ ,$$

a formula obtained by Watson (1984) for an ARMA model case with $J=(-\infty,\infty)$, under Bell's Assumption A.

To illustrate the generality of the results obtained here and in the previous section, let us note that the relations (4.1), (4.2) and (3.3) transform into similar formulas if e^{α}_{t} is replaced by a nonstationary moving average $\bar{e}^{\alpha}_{t} = \Sigma_i\, c_{t,i}e^{\alpha}_{t-i}$. Thus, if the coefficients $c_{t,i}$ are known, then by making straightforward changes like those described above to the formulas of section 3, we can solve the analogues of (1.6) and (1.7) with n^{α}_{t} replaced by $\bar{n}^{\alpha}_{t} = n_t+\bar{e}^{\alpha}_{t}$, $m\leqq\alpha\leqq M$.

5. BELL'S ASSUMPTION A AND THE ASSUMPTIONS (3.2) AND (3.3)

In this section, we show that, in conjunction with other standard assumptions, the initializing Assumption A of Bell (1984) and our (4.1) imply that (3.2) and (3.3) hold, with functions f(r,t) and g(r,t) which are easily determined. This shows that our analyses of (1.6) and (1.7) are applicable when this assumption is made.

We now assume that the random variables x_t, n_t and e^{α}_{t}, $m\leqq\alpha\leqq M$, are defined for all $t = 0,\pm1, \ldots$, each with mean zero and finite variance, where the series e^{α}_{t} is *stationary* and satisfies (4.1) with $d(r,t) = d_e(r-t)$ for some symmetric positive definite function $d_e(\cdot)$. Motivated by the situation described in Appendix 1, we shall also assume the existence of minimal-degree backshift-operator polynomials $\delta_n(B)$ and $\delta_s(B)$, *not depending on* α, which transform the series $n^{\alpha}_{t} = n_t+e^{\alpha}_{t}$ and $s^{\alpha}_{t} = x_t-n^{\alpha}_{t}$ into *uncorrelated* mean-zero stationary series, $u^{\alpha}_{t} = \delta_n(B)n^{\alpha}_{t}$ and $v^{\alpha}_{t} = \delta_s(B)s^{\alpha}_{t}$, $m\leqq\alpha\leqq M$. If $\delta_c(B)$ denotes the greatest common divisor of $\delta_n(B)$ and $\delta_s(B)$, then quite generally (Findley, 1985), the corresponding minimal degree polynomial transformation to stationarity for x_t is $\delta_x(B) = \delta_n(B)\delta_s(B)/\delta_c(B)$, which has degree $d(x) = d(n)+d(s)-d(c)$, using an obvious notation. For $m\leqq\alpha\leqq M$, we clearly have

$$u_t^\alpha = u_t^m + \delta_n(B)e_t^\alpha \tag{5.1}$$

$$v_t^\alpha = v_t^M + \delta_s(B)\{e_t^M - e_t^\alpha\} \tag{5.2}$$

and

$$\begin{aligned}\delta_x(B)x_t &= \{\delta_s(B)/\delta_c(B)\}u_t^\alpha + \{\delta_n(B)/\delta_c(B)\}v_t^\alpha \\ &= \{\delta_s(B)/\delta_c(B)\}u_t^m + \delta_x(B)e_t^\alpha + \{\delta_n(B)/\delta_c(B)\}v_t^\alpha .\end{aligned} \tag{5.3}$$

The two series in the first decomposition in (5.3) are uncorrelated because this is true of the series u_t^α and v_t^α, but, with the motivation of Appendix 1, we shall make the slighty stronger assumption that the three series in the final decomposition in (5.3) are uncorrelated. Using the inversion procedures described in Findley (1985), it is easy to see that this is equivalent to the following assumption.

(5.4) *For any* α *in* $[m,M]$, *the series* $\{u_t^m\}$, $\{e_t^\alpha\}$ *and* $\{v_t^\alpha\}$ *are uncorrelated.*

We have used curly brackets in (5.4) to emphasize that each observation of one series is assumed to be uncorrelated with all observations of the two other series. The formulation (5.4) is convenient when a state space representation permitting the use of the Kalman Filter algorithm is sought. as the example in the next section illustrates.

Suppose that $\delta_x(B) = 1 - \delta_1 B - \dots - \delta_{d(x)}B^{d(x)}$ and that $1, \xi_1, \xi_2, \dots$ are the coefficients of the formal expansion $[\delta_x(B)]^{-1} = 1 + \xi_1 B + \xi_2 B^2 + \dots$. Define the series $w_t = \delta_x(B)x_t$. Then

$$x_t = \delta_1 x_{t-1} + \dots + \delta_{d(x)}x_{t-d(x)} + w_t \tag{5.5}$$

holds for all t, from which it follows that the series $\{x_t\}$ determines and is completely determined by the series $\{w_t\}$ and a set of d(x) starting values, say, $x_1, \dots, x_{d(x)}$. From (5.5) it follows that, for $t>d(x)$,

$$x_t = \sum_{i=1}^{d(x)} a_{i,t}x_i + \sum_{j=0}^{t-d(x)-1} \xi_j w_{t-j} , \tag{5.6}$$

with a similar equation holding for $t \leq 0$, see Bell (1984) for details. From (5.6) and (5.4) we obtain, for example, that for any q,r and any α in [m,M],

$$\mathrm{cov}(w_q,\ e_r^\alpha) = -(\alpha - m) \sum_{j=0}^{d(x)} \delta_j\ d_e(q-j-r)\ . \tag{5.7}$$

Thus (3.3) will hold for all values of $t-j$ if it holds for $t-j = 1, \ldots, d(x)$. Also, if this happens, then (3.2) will hold : Indeed, by an argument analogous to that given above, (3.2) will hold for all values of t if it holds for the initializing values $n_1, \ldots, n_{d(n)}$, and equation (4.1) of Bell (1984) and our (5.1) and (5.2) show that these values can be written as linear combinations of x_i, u_i^m, e_i^α, v_i^α, $i=1, \ldots, d(x)$ with coefficients which are independent of α. Clearly, therefore, (3.2) and (3.3) will hold if, for example, for all values of t and α,

$$\mathrm{cov}(x_i,\ e_t^\alpha) = 0\ ,\quad 1 \leq i \leq d(x)\ . \tag{5.8}$$

We will now show that (5.8) is implied by the fruitful starting value specification described as Assumption A in Bell (1984), which has the following formulation in our context.

(5.9) *Bell's Assumption A : The starting values* $x_1, \ldots, x_{d(x)}$ *are uncorrelated with the series* $\{u_t^\alpha\}$ *and* $\{v_t^\alpha\}$, $m \leq \alpha \leq M$, *and* (5.4) *holds* .

In fact, it follows immediately from (5.1) and (5.2) (first choose $\alpha = m$) and (5.9) that $x_1, \ldots, x_{d(x)}$ are uncorrelated with all mean square limits of linear combinations of $\delta_n(B)e_t^\alpha$, $t = 0, \pm 1, \ldots$. Since e_t^α , $t = 0, \pm, \ldots$ can be obtained from such limits via the inverse filter $\delta_n^{-1}(B)$, see Findley (1985), the condition (5.8) is a consequence of (5.9), as are thus also (3.2) (and (4.2)) and (3.3) . This is what we wanted to show.

6. A SIMPLE EXAMPLE

In this section, we use a simple ARMA model example to illustrate what is involved in carrying out the solution procedure of subsection (3.2) under assumption (5.9) in the usual seasonal ARMA model context, and to demonstrate some consequences, including two rather unexpected ones, of this assumption.

Consider a "semiannual" time series x_t satisfying the seasonal model

$$(1-B^2)x_t = \varepsilon_t - \theta\varepsilon_{t-1} \quad (|\theta|<1) \tag{6.1}$$

with initializing values x_1 and x_2, where ε_t is a white noise proccess with variance $E\varepsilon_t^2 = 1$. The pseudospectrum $f_x(\lambda)$ of x_t has the family of pseudospectral decompositions

$$\frac{|1-\theta e^{i\lambda}|^2}{|1-e^{i2\lambda}|^2} = \left\{ \frac{1}{4(1-\theta)^2}\frac{1}{|1-e^{i\lambda}|^2} + \alpha \right\} + \left\{ \frac{1}{4(1+\theta)^2}\frac{1}{|1+e^{i\lambda}|^2} - \alpha \right\}$$

$$= f_n^\alpha(\lambda) + f_s^\alpha(\lambda)\ , \ m \leqq \alpha \leqq M\ , \tag{6.2}$$

where $m = -\min_\lambda f_n^0(\lambda)$ and $M = \min_\lambda f_s^0(\lambda)$. Since $\max_\lambda|1\pm e^{i\lambda}|^2 = \max_\lambda\{2\pm 2\cos\lambda\}=4$, it follows that $m = -(1-\theta)^{-2}/16$ and $M = (1+\theta)^{-2}/16$. The component pseudospectra $f_n^\alpha(\lambda)$ and $f_s^\alpha(\lambda)$ are those of processes n_t^α and s_t^α satisfying

$$\begin{aligned}(1-B)n_t^\alpha &= u_t^\alpha \\ (1+B)s_t^\alpha &= v_t^\alpha\ ,\end{aligned} \tag{6.3}$$

where u_t^α and v_t^α are white noise processes if $\alpha=0$ and moving average processes of order 1 otherwise, for $m \leqq \alpha \leqq M$. For example, u_t^α has the spectral density

$$\begin{aligned}f_u^\alpha(\lambda) &= \{4(1-\theta)^2\}^{-1} + \alpha|1-e^{i\lambda}|^2 \\ &= \sigma_u^2(\alpha)\ |1-\psi^\alpha e^{i\lambda}|^2\end{aligned} \tag{6.4}$$

with $\sigma_u^2(\alpha)$ and $\psi^\alpha(|\psi^\alpha| \leqq 1)$ determined by the second equation in (6.4), which is equivalent to the pair of nonlinear equations

$$\{4(1-\theta)^2\}^{-1} + 2\alpha = \sigma_u^2(\alpha\ \{1+(\psi^\alpha)^2\} \tag{6.5}$$

$$\alpha = \sigma_u^2(\alpha)\psi^\alpha .$$

In particular, $\psi^m = -1$ and $\sigma_u^2(m) = -m$, so that, since $n_t = n_t^m$,

$$(1-B)n_t = a_t^m + a_{t-1}^m , \tag{6.6}$$

where a_t^m is a white noise process with variance $-m$. Thus, for the decomposition model

$$x_t = n_t + e_t^0 + s_t^0 ,$$

with white noise process e_t^0, corresponding to the pseudospectrum decomposition

$$f_x(\lambda) = f_n^m(\lambda) + (-m) + f_s^0(\lambda) ,$$

we have the state space representation

$$\begin{bmatrix} n_t \\ a_t^m \\ s_t^0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} n_{t-1} \\ a_{t-1}^m \\ s_{t-1}^0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_t^m \\ v_t^0 \end{bmatrix},$$

$$x_t = [\,1 \quad 0 \quad 1\,] \begin{bmatrix} n_t \\ a_{t-1}^m \\ s_t^0 \end{bmatrix} + e_t^0 \quad . \tag{6.7}$$

The Kalman Filter and related smoothing algorithms are applicable to (6.7) for $t \geqq t_1$ provided that the series $\{e_t^0\}$ is uncorrelated with the series $\{a_t^m\}$ and $\{v_t^0\}$ (guaranteed by (5.4)) and that, for times $t \geqq t_1+1$, the variables a_t^m, v_t^0 and e_t^0 are uncorrelated with n_{t_1} and $s_{t_1}^0$. This condition is satisfied under (5.9) if the designation of the observation time origin is chosen in such a way that $t_1 \geqq 2$. (Recall our convention that $t=1$ denotes the initializing time for the ARMA equation). To see this, note that since $n_2^\alpha = n_1^\alpha + u_2^\alpha$ and $s_2^\alpha = -s_1^\alpha + v_2^\alpha$, we have

$$x_1 = n_1^\alpha + s_1^\alpha$$

$$x_2 = n_2^\alpha + s_2^\alpha = n_1^\alpha - s_1^\alpha + u_2^\alpha + v_2^\alpha \;, \tag{6.8}$$

whence

$$\begin{aligned} n_i^\alpha &= \{x_1 + x_2 + (-1)^i u_2^\alpha - v_2^\alpha\}/2 \;,\; i = 1,2 \\ s_i^\alpha &= (-1)^i \; \{x_2 - x_1 - u_2^\alpha + (-1)^i v_2^\alpha\}/2, \; i = 1,2 \;. \end{aligned} \tag{6.9}$$

Since

$$\begin{aligned} u_t^\alpha &= u_t^m + (1-B)\; e_t^\alpha \\ v_t^\alpha &= v_t^M + (1+B)\; \{e_t^M - e_t^\alpha\} \;, \end{aligned} \tag{6.10}$$

it follows from (6.9) that a_t^m, v_t^0 and e_t^m are uncorrelated with n_2 and s_2^0 for $t \geqq 3$.

The Kalman Filter algorithm for (6.7) initialized at $t=t_1$ requires the variance ($= -m$) of e_t^0, the covariance matrix $\Sigma(a_t^m, v_t^0)$ of the noise vector $[a_t^m \; v_t^0]^T$ and the covariance matrix $\Sigma(n_{t_1}, a_{t_1}^m, s_{t_1}^0)$ of the initializing state vector. From (6.2) and (6.5)

$$\Sigma(a_t^m, v_t^0) = \begin{bmatrix} -m & 0 \\ 0 & 4M \end{bmatrix} . \tag{6.11}$$

For the computation of $\Sigma(n_{t_1}, a_{t_1}^m, s_{t_1}^0)$, it is important to observe from (6.9) and (6.10) that n_i^α, s_i^α,$i = 1,2$ are correlated with e_i^α, $i = 1,2$. Rather unexpectedly, this happens in such a way that $\mathbf{var}(n_i^\alpha)$, $\mathbf{var}(s_i^\alpha)$ and $\mathbf{cov}(n_i^\alpha, s_i^\alpha)$ do not depend on α. (Thus, in particular, the addition of e_i^α to n_i does not increase the variance.) Indeed, the contribution of u_2^α and v_2^α to these quantities is $\pm(E\{u_2^\alpha\}^2 + E\{v_2^\alpha\}^2)/4 = \pm(M-m)$. Consequently, for $t_1=2$ we have

$$\Sigma(n_2,\, a_2^m,\, s_2^0) = \begin{bmatrix} \frac{1}{4}E\{x_1+x_2\}^2+M-m & . & . \\ -m/2 & -m & . \\ \frac{1}{4}\{Ex_2^2-Ex_1^2\}-M+m & 0 & \frac{1}{4}E\{x_2-x_1\}^2+M-m \end{bmatrix} . \qquad (6.12)$$

After these preliminaries, the fixed interval smoothing algorithm can be applied as indicated in subsection (3.2) to solve (1.6). In order to illustrate explicitly the dependence of the solution $\beta^*(t)$ of (1.6) on the parameter θ in (6.1), we now turn to the asymptotic case $J = (-\infty,\infty)$, where, under (5.9), the filter coefficients for $\hat{n}_t^\alpha$ are determined by

$$\frac{f_n^\alpha(\lambda)}{f_x(\lambda)} = |1-\theta e^{i\lambda}|^{-2}\ [\{4(1-\theta)^2\}^{-1}|1+e^{i\lambda}|^2+\alpha|1-e^{i2\lambda}|^2]$$

$$= \Sigma_{j=-\infty}^{\infty}\ h_j^\alpha\ e^{ij\alpha}\ ,$$

where

$$h_j^\alpha = (1-\theta^2)^{-1} \left\{ \frac{\theta^{|j+1|}+2\theta^{|j|}+\theta^{|j-1|}}{4(1-\theta)^2} - \alpha(\theta^{|j+2|}- 2\theta^{|j|}+\theta^{|j-2|}) \right\} . \qquad (6.13)$$

We will determine $\beta_0(t)$ with the aid of (3.14) and (3.15). For this example, $d(t,t)=1$, and $f(t,t)$ and $g(r,t)$ are easy to calculate. Since $\mathbf{cov}(u_t^m,\ e_t^\alpha) = 0$ and $v_2^m = v_2^\alpha-(1+B)e_t^\alpha$, it follows from (6.3) and (6.9) that

$$\mathbf{cov}(n_t,\ e_t^\alpha) = \mathbf{cov}(n_1,\ e_t^\alpha) = -\ \mathbf{cov}(e_2^\alpha+e_1^\alpha,\ e_t^\alpha)/2\ .$$

Hence

$$f(t,t) = \begin{cases} -1/2\ , & t=1,2 \\ 0\ \ , & \text{otherwise} \end{cases} . \qquad (6.14)$$

The formula for $g(r,t) = (\alpha-m)^{-1}\ \mathbf{cov}(x_r,\ e_t^\alpha)$ is slightly more complex. Set $w_t = (1-B^2)x_t$. Then, from (6.1), we have

$$x_r = \begin{cases} x_2 + \sum_{k=2}^{n} w_{2k} , & r=2n, \quad n \geqq 2 \\ x_1 + \sum_{k=2}^{n} w_{2k-1} , & r=2n-1, \quad n \geqq 2 \\ x_2 - \sum_{k=n+1}^{1} w_{2k} , & r=2n, \quad n \leqq 0 \\ x_1 - \sum_{k=n+1}^{1} w_{2k-1} , & r=2n-1, \quad n \leqq 0 \end{cases} \quad (6.15)$$

Since $w_t = (1+B)u_t^m+(1-B^2)e_t^\alpha+(1-B)v_t^\alpha$ and $\Sigma_{k=2}^n (1-B^2)e_{2k}^\alpha = e_{2n}^\alpha - e_2^\alpha$, etc., it follows from (5.9) and (6.15) that

$$g(r,t) = \begin{cases} 0 & \text{all } t , & r = 1,2 \\ 1 & t=r , & r \neq 1,2 \\ -1 & t=1 , & r \text{ odd} , \ \neq 1 \\ -1 & t=2 , & r \text{ even} , \ \neq 2 \\ 0 & \text{otherwise} . \end{cases} \quad (6.16)$$

For $t\neq 1,2$, therefore, (3.16) applies and, from (6.13),

$$C_t(\beta) = 1 - 4\beta - (1-\theta)^{-3} ,$$

so that for $t\neq 1,2$

$$\beta_0(t) = \{1 - (1-\theta)^{-3}\}/4 . \quad (6.17)$$

Thus, for example, $\beta_0(t\neq 1,2)\in[m,M]$, and $\beta^*(t\neq 1,2)=\beta_0(t\neq 1,2)$ if and only if

$$-(1+\theta)^2(1-\theta) \leqq \{(1-\theta)^3(1+\theta)^2-(1+\theta)^2\} \leqq (1-\theta)^3$$

whose solution can be numerically approximated to any desired accuracy. From (6.17) we are led to the following numerically approximate solution of (1.6) ,

$$\beta^*(t\neq 1,2) = \begin{cases} m \;, & \theta \in \quad [0.083,\ 1] \\ \beta_0(t) \;, & \theta \in \quad [-1,\ -0.352] \cup [-0.154,\ 0.083] \\ M \;, & \theta \in \quad [-0.352,\ -0.154] \end{cases} .$$

For the case when $t=1,2$, we obtain from (6.14) and (6.16) that $C_t(\beta) = 2\Sigma_{j\neq 0}\, h_{2j}^{\alpha}$, so that, from (6.13) ,

$$C_t(\beta) = \qquad \theta(1-\theta)^{-4} - 4\alpha \;,$$

which reveals that

$$\beta_0(t=1,2) = \frac{1}{4}\, \theta(1-\theta)^{-4} \;.$$

Applying (2.6), it can be verified that, to the rounded value,

$$\beta^*(t=1,2) = \begin{cases} M \;, & \theta \in [0.435,\ 1) \\ \beta_0 \;, & \theta \in (-1,\ 0.435] \end{cases}$$

which is clearly different from the values $\beta^*(t\neq 1,2)$ obtained above from Watson's formula (3.16).

To return to the discussion of the method of subsection 3.2, the example discussed above illustrates that once a nonlinear system of equations, exemplified by (6.5), has been solved ("spectral factorization"), then the determination of initializing covariance matrices like (6.11) and (6.12) under (5.9) involves only straightforward calculations. Bell (1984) provides the needed generalization of (6.9). If, instead of (5.9), it is assumed that the component series $\{n_t\}$, $\{e_t^{\alpha}\}$ and $\{s_t^{\alpha}\}$ are uncorrelated with one another, then the determination of the initializing covariance matrices is substantially more complex, due, in part, to the requirement that the covariance structure of x_t defined by (6.7) should not depend on α, see Findley (1985b). We observe, finally, that there are good numerical algorithms for performing the spectral factorization, see Anderson and Gartland (1985).

APPENDIX 1 : SOURCES OF (1.3) – (1.4) WITH MOVING AVERAGE e_t^α

The model–based approaches to seasonal adjustment referred to in section 1 typically have the following ingredients. There is a minimal–degree backshift–operator polynomial transformation $\delta_x(B)$, whose roots are on the unit circle, such that $w_t = \delta_x(B)x_t$ is a mean–zero stationary ARMA(p,q) process, with spectral density $f_w(\lambda)$, say, whose autoregressive polynomal will be denoted by $\varphi(B)$. Further, $\delta_x(B)$ admits a factorization $\delta_x(B) = \delta_n(B)\delta_s(B)$ into components associated with the nonseasonal, respectively, seasonal components of x_t. For example, if $\delta_x(B) = 1-B^{12}$, then $\delta_n(B) = 1-B$ and $\delta_s(B) = 1+B+ \ldots +B^{11}$. There usually exists a partial fraction decomposition of the pseudospectrum $f_x(\lambda) = f_w(\lambda)/|\delta_x(e^{i\lambda})|^2$ of x_t of the form

$$\frac{f_w(\lambda)}{|\delta_x(e^{i\lambda})|^2} = \frac{ma_u(\lambda)}{|\delta_n(e^{i\lambda})\varphi(e^{i\lambda})|^2} + \frac{ma_v(\lambda)}{|\delta_s(e^{i\lambda})|^2} + ma_0(\lambda)$$

$$= f_n^0(\lambda) + f_s^0(\lambda) + ma_0(\lambda)\ , \qquad \text{(A1.1)}$$

where $ma_u(\lambda)$, $ma_v(\lambda)$ and $ma_0(\lambda)$ are spectral densities of moving average processes of orders $q(u) \leqq d(n) + p-1$, $q(v) \leqq d(s)-1$ and $q(0) = q-p-d(x)$, respectively. (We set $ma_0(\lambda) \equiv 0$ if this last number is negative. Order 0 refers to a white noise process.) Frequently there are additional decompositions of $f_x(\lambda)$ having the same denominators, and these denominators are often the only really compelling feature of the decomposition. One possibility is illustrated in the example of section 6. Another is that $ma_0(\lambda)$, if nonzero, can be decomposed into components which are assigned to the other two terms. Or, more generally, suppose for example, that $ma_u(\lambda)$ is always positive on $[-\pi,\pi]$. Let $ma_1(\lambda)$ be the spectral density of a moving average process, of any order $q(1)\geqq 0$, with the property that $ma_u(\lambda) - ma_1(\lambda)\ |\delta_n(e^{i\lambda})\varphi(e^{i\lambda})|^2$ is always non–negative, and so is a moving average spectral density. Then there is an associated family of pseudospectrum decompositions

$$f_x(\lambda) = \{f_n(\lambda)-ma_1(\lambda)+\alpha[ma_0(\lambda)+ma_1(\lambda)]\} + \{f_s(\lambda)+(1-\alpha)[ma_0(\lambda)+ma_1(\lambda)]\}$$

$$= \frac{ma_u^\alpha(\lambda)}{|\delta_n(e^{i\lambda})\varphi(e^{i\lambda})|^2} + \frac{ma_v^\alpha(\lambda)}{|\delta_s(e^{i\lambda})|^2}\ , \quad m\leqq\alpha\leqq M \qquad \text{(A1.2)}$$

associated with decompositons of x_t of the form (1.3) – (1.4), where m could be defined as the smallest number for which the first expression in curly brackets in (A1.2) is always nonnegative (so $m \leqq 0$) and M could be chosen to be the largest positive number for which the other expression in curly brackets is always nonnegative.

In this example, the "minimal" process n_t has pseudosprectrum

$$f_n(\lambda) + (m-1)\{ma_1(\lambda)\} + m\{ma_0(\lambda)\}$$

and the process e_t^{α} $(\alpha > m)$ is a moving average process of order not exceeding max $\{q(0),q(1)\}$ with spectral density $(\alpha - m)\{ma_0(\lambda)+ma_1(\lambda)\}$.

A somewhat different situation motivating the use of an $ma_1(\lambda)$ term is the following. Sometimes the partial fraction decompostion of $f_x(\lambda)$ leads to a form of (A1.1) which is improper in the sense that one (or more) of the numerators is negative for some values of λ. If, for example, $f_n(\lambda)+f_s(\lambda)$ is nonnegative, but $f_s(\lambda)$ is negative for some λ, one could seek an $ma_1(\lambda)$ with the property that both $f_n(\lambda)-ma_1(\lambda)$ and $f_s(\lambda)+ma_1(\lambda)$ are nonnegative.

APPENDIX 2 : PROOFS OF (2.5) – (2.7)

Differentiating $F(\alpha,\beta)$ with respect to β, we obtain

$$F_{\beta}(\alpha,\beta) = 2D(\beta-\alpha) \, ,$$

which reveals that for each fixed α, the function $\beta \to F(\alpha,\beta)$ is uniquely minimized at $\beta = \alpha$, establishing (2.5). The derivative of the function $\alpha \to F(\alpha,\alpha)$ is given by

$$F_{\alpha}(\alpha,\alpha) = C - 2D(\alpha - m) \tag{A2.1}$$

so that $F(\alpha,\alpha)$ is uniquely maximized at $\alpha = \beta_0$. Since $F(\alpha,\beta_0)$ is independent of α, we thus have

$$F(\alpha,\alpha) < F(\alpha,\beta_0) = F(\beta_0,\beta_0) < F(\beta_0,\beta) \tag{A2.2}$$

for all $\alpha \neq \beta_0 \neq \beta$. As a consequence, we obtain

$$\max_\alpha F(\alpha,\alpha) = F(\beta_0,\beta_0) \leqq \min_\beta \max_\alpha F(\alpha,\beta) .$$

Since the reversed inequality is obvious, the assertions of (2.6) and (2.7) follow for the case ((i)) in which $m \leqq \beta_0 \leqq M$, the uniqueness assertion being a consequence of the strict inequality in (A2.2).

In case (ii), $C-2D(\beta-m)$ is negative if $m \leqq \beta \leqq M$, so for fixed β, $F(\alpha,\beta)$ is maximized at $\alpha=m$. Since $F(m,\beta)$ is uniquely minimized at $\beta=m$, the assertion of (2.6) follows. The assertions of (2.7) follow from (A2.1) or by inspection.

Case (iii) is handled similarly to case (ii).

REFERENCES

ANDERSON, DAVID H. and GARTLAND, EUGENE C. (1985). "The Numerical Solution of a Nonlinear System Arising in Time Series Analysis." *SIAM J. Sci. Stat. Comput.* 6, 349–364.

ANDERSON, BRIAN D.O. and MOORE, JOHN B. (1979). *Optimal Filtering.* Englewood Cliffs : Prentice Hall.

BELL, WILLIAM (1984). "Signal Extraction for Nonstationary Time Series." *Annals of Statistics* 12, 646–664.

BURMAN, J. PETER (1980). "Seasonal Adjustment by Signal Extraction." *JRSS Series A.* 143, 321–337.

FINDLEY, DAVID F. (1985a). "On Backshift-Operator Polynomial Transformations to Stationarity for Nonstationary Series and Their Aggregates." *Communications in Statistics,* A(14), 49–62.

FINDLEY, DAVID F. (1985b). "A State-Space Smoother Implementation of a Generalization of M. Watson's Minimax Approach to Model-Based Seasonal Adjustment." *Proceedings of the Business and Economics Section of the American Statistical Association.*

HILLMER, STEVEN C., BELL, WILLIAM and TIAO, GEORGE C. (1983). "Modeling Considerations in the Seasonal Adjustment of Economic Time Series." In *Applied Time Series Analysis of Economic Data* (Editor, Arnold Zellner), Washington, D.C.: U.S. Department of Commerce, Bureau of Census.

PARZEN, E. (1984). *Time Series Analysis of Irregularly Spaced Data.* Lecture Notes in Statistics, vol. 25. Heidelberg : Springer Verlag.

WATSON, MARK W. (1984). "Uncertainty in Model-Based Seasonal Adjustment Procedures: Bounds on Mean Square Error und Construction of Minimax Filters." Harvard Institute of Economic Research Discussion Paper Number 1071.

Footnote (page 4)

1) It is implicit in the optimality and convergence assumptions of the papers reference above that the random variables x_t, n_t^α and s_t^α have finite variances (increasing to ∞ as $t \to \infty$) : This is merely the assumption that the random variables initializing the ARMA difference equation for x_t have finite variance, see Bell (1984), especially Bell's equation (3.1). Thus (1.5) and the mean square expressions in (1.6) and (1.7) below are finite.