BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/33

A GENERAL ANALYSIS OF WATSON'S MINIMAX PROCEDURE FOR COMPONENT MODEL SELECTION IN NON-STATIONARY ARMA MODEL-BASED SEASONAL ADJUSTMENT

by

David F. Findley
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

Recommended by: Nash J. Monsour

Report completed: November 29, 1984

Report issued: December 10, 1984

On Watson's Minimax Procedure for Component Model Selection Model-Based Seasonal Adjustment.

David F. Findley
Bureau of the Census

1. Introduction

In applications of the Hillmer-Tiao model-based signal extraction approach to seasonal adjustment (see Hillmer, Bell and Tiao (1983) or Burman (1980)), there usually is a white noise component which must be divided among the seasonal and nonseasonal components, or assigned wholly to one of them (see the example in section 4 below). This assignment determines the covariance structure which must be specified before the optimal estimates, or, essentially equivalently, the filters used to obtain them, can be determined. Recently, Watson (1984) proposed the use of a minimax criterion related to mean square component-estimation-error for making this assignment. His approach is quite attractive, due in part to additional appealing properties he demonstrates for some of the solutions to his minimax problem.

Watson gives a detailed discussion of his approach only for the situation in which the coefficients of the optimal signal extraction filter are obtained from pseudo-spectrum ratios, in analogy with the classical Wiener-Kolmogoroff theory, as under Assumption A of Bell (1984) with bi-infinite data. It is not clear from his arguments whether his results generalize to other situations, or if the solution to the minimax problem can be determined in situations where the optimal extraction filters are not explicitly available, such as when a Kalman smoothing algorithm is used.

In section 2, we give a "geometrical" development of his results which makes no use of filters or models but only of the covariance properties of least mean square estimates. This shows that Watson's conclusions have very general validity. Sections 3 and 4 discuss and exemplify how one could use a Kalman smoother algorithm to obtain the minimax solutions. The analysis given of the example seems to be the first to show what is required in order for the initial covariance specification to precisely replicate the structure implicit in the Hillmer-Tiao component decomposition. Section 5 gives a general analysis of Watson's second minimax criterion and its solutions. This second criterion is more appropriate when estimates of nonseasonal period-to-period-changes are the quantities of primary interest.

2. A General Analysis of Watson's First Minimax Criterion.

Let x_t be a time series with mean 0, $Ex_t=0$, and finite but possibly time-varying variance, $\mathrm{var}(x_t) = Ex_t^2 < \infty$. For example, x_t could be a seasonal ARIMA process whose initial values have mean zero and finite variance. We consider the situation in which x_t admits a family of decompositions

$$x_t = s_t^\gamma + n_t^\gamma \ , \quad \gamma^\ell \leq \gamma \leq \gamma^u$$

having the property that for any $\gamma,\gamma' \varepsilon [\gamma^\ell,\gamma^u]$ with $\gamma' < \gamma$, the series

$$e_t^{\gamma,\gamma'} =_{\mathrm{def}} n_t^{\gamma'} - n_t^\gamma = s_t^\gamma - s_t^{\gamma'}$$

is a white noise process with $\mathrm{var}(e_t^{\gamma,\gamma'}) = \gamma - \gamma'$ which is uncorrelated with the processes n_t^{γ} and $s_t^{\gamma'}$,

$$\mathrm{cov}(e_t^{\gamma,\gamma'}, s_u^{\gamma'}) = \mathrm{cov}(e_t^{\gamma,\gamma'}, n_u^{\gamma}) = 0$$

for all t,u. We do not need to assume that the component series $s_t^{\gamma'}$ and n_t^{γ} are uncorrelated.

Throughout the remainder of this section it will be convenient to hold t fixed and assume that the observed random variables are x_{t-i}, $-m \leq i \leq n$ for some possibly infinite $m,n \geq 0$. (Set $x_{-\infty} = x_{\infty} = 0$.) Let OBS(=OBS(t,m,n)) denote the linear space consisting of all finite linear combinations of x_{t-i}, $-m \leq i \leq n$ (and their mean square limits if m or n 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. It follows from this that if we have two such random variables y and z, then $\hat{y} \pm \hat{z}$ is the least mean square approximation in OBS to $y \pm z$. Also, note that if $\mathrm{cov}(y, x_{t-i}) = \mathrm{cov}(z, x_{t-i})$, $-m \leq i \leq n$, then $y - z$ is uncorrelated with OBS, so that $\hat{y} - \hat{z} = 0$, i.e., $\hat{y} = \hat{z}$.

For $\gamma^i, \gamma^j \in [\gamma^{\ell}, \gamma^u]$, let $a_t^{\gamma^i,\gamma^j}$ denote the error when $\hat{n}_t^{\gamma^j}$ is used to approximate $n_t^{\gamma^i}$, i.e.

$$a_t^{\gamma^i,\gamma^j} = n_t^{\gamma^i} - \hat{n}_t^{\gamma^j} \quad ,$$

and denote the mean square of $a_t^{\gamma^i,\gamma^j}$ by $\mathrm{ms}[a_t^{\gamma^i,\gamma^j}]$ as in Watson (1984).

To choose among the possible estimates $\hat{n}_t^{\gamma^j}$, $\gamma^j \varepsilon [\gamma^\ell, \gamma^u]$ of the non-seasonal components of x_t, Watson suggests that the criterion

$$\min_{\gamma^j} \max_{\gamma^i} ms[a_t^{\gamma^i, \gamma^j}] \tag{2.1}$$

should be used.

We begin our analysis of (2.1) by recalling that for each $\gamma \varepsilon (\gamma^\ell, \gamma^u]$ there is a decomposition of $n_t^{\gamma^\ell}$ into uncorrelated components,

$$n_t^{\gamma^\ell} = n_t^{\gamma} + e_t^{\gamma, \gamma^\ell} . \tag{2.2}$$

To simplify the exposition, we shall initially consider the decomposition of x_t given by $x_t = s_t^{\gamma^\ell} + e_t^{\gamma, \gamma^\ell} + n_t^{\gamma}$.

Observe that for any $\gamma \varepsilon (\gamma^\ell, \gamma^u]$ and $-m<i<n$, we have

$$\operatorname{cov}((\gamma-\gamma^\ell)^{-1} e_t^{\gamma,\gamma^\ell}, x_{t-i}) = (\gamma-\gamma^\ell)^{-1} \operatorname{cov}(e_t^{\gamma,\gamma^\ell}, e_{t-i}^{\gamma,\gamma^\ell}) = \begin{cases} 1, i=0 \\ 0, i\neq 0 \end{cases} \tag{2.3}$$

It follows that for any $\hat{y}_t \varepsilon OBS$, $(\gamma-\gamma^\ell)^{-1} \operatorname{cov}(e_t^{\gamma,\gamma^\ell}, \hat{y}_t)$ is independent of γ and that if

$$\hat{y}_t = \sum_{-m<i<n} v_i x_{t-i} \tag{2.4}$$

(assuming convergence in mean square if m or n is infinite), then

$$(\gamma-\gamma^{\ell})^{-1}\mathrm{cov}(\hat{y}_t, e_{t-i}^{\gamma,\gamma^{\ell}}) = v_i \quad , \ -m<i<n \ . \tag{2.5}$$

It also follows that $\tilde{e}_t = (\gamma-\gamma^{\ell})^{-1}\hat{e}_t^{\gamma,\gamma^{\ell}}$, the least mean square approximation in OBS of $(\gamma-\gamma^{\ell})^{-1}e_t^{\gamma,\gamma^{\ell}}$, does not depend on γ. Set $h_0 = \mathrm{var}(\tilde{e}_t)$. For any $\gamma^i, \gamma^j \varepsilon [\gamma^{\ell}, \gamma^u]$, we clearly have

$$\mathrm{cov}(\hat{e}_t^{\gamma^i,\gamma^{\ell}}, \hat{e}_t^{\gamma^j,\gamma^{\ell}}) = (\gamma^i-\gamma^{\ell})(\gamma^j-\gamma^{\ell})h_0 \ . \tag{2.6}$$

We note, too, that for any $y_t \varepsilon$ OBS, since

$$n_t^{\gamma^{\ell}} - \hat{y}_t = n_t^{\gamma^i} - \hat{y}_t + e_t^{\gamma^i,\gamma^{\ell}}$$

and since $e_t^{\gamma^i,\gamma^{\ell}}$ and $e_t^{\gamma^i,\gamma^{\ell}} - \hat{e}_t^{\gamma^i,\gamma^{\ell}}$ are uncorrelated with $n_t^{\gamma^i}$ and $\hat{y}_t$, respectively, we further have

$$E\{n_t^{\gamma^{\ell}} - \hat{y}_t\}^2 = E\{n_t^{\gamma^i} - \hat{y}_t\}^2 + (\gamma^i - \gamma^{\ell}) - 2\ \mathrm{cov}(e_t^{\gamma^i,\gamma^{\ell}}, \hat{y}_t)$$

$$= E\{n_t^{\gamma^i} - \hat{y}_t\}^2 + (\gamma^i - \gamma^{\ell})\{1 - 2\ \mathrm{cov}(\tilde{e}_t, \hat{y}_t)\} \ . \tag{2.7}$$

By the preceding discussion, the coefficient of $(\gamma^i - \gamma^\ell)$ in (2.7) is independent of γ^i and is equal to $1 - 2v_0$ if $\hat{y}_t$ is given by (2.4).

Setting $\hat{y}_t = \hat{n}_t^{\gamma^j}$ in (2.7) and defining $h_0^{\gamma^j} = \mathrm{cov}(\tilde{e}_t, \hat{n}_t^{\gamma^j})$, we obtain

$$\mathrm{ms}[a_t^{\gamma^\ell}, \gamma^j] = \mathrm{ms}[a_t^{\gamma^i}, \gamma^j] + (\gamma^i - \gamma^\ell)\{1 - 2h_0^{\gamma^j}\} \tag{2.8}$$

On the other hand, since

$$n_t^{\gamma^\ell} - \hat{n}_t^{\gamma^j} = \{n_t^{\gamma^\ell} - \hat{n}_t^{\gamma^\ell}\} + \{\hat{n}_t^{\gamma^\ell} - \hat{n}_t^{\gamma^j}\}$$

$$= \{n_t^{\gamma^\ell} - \hat{n}_t^{\gamma^\ell}\} + \hat{e}_t^{\gamma^j, \gamma^\ell} \quad ,$$

we also have from (2.6) that

$$\mathrm{ms}[a_t^{\gamma^\ell}, \gamma^j] = \mathrm{ms}[a_t^{\gamma^\ell}, \gamma^\ell] + (\gamma^j - \gamma^\ell)^2 h_0 \ .$$

Substituting this expression into (2.8) and rearranging, we arrive at the fundamental formula

$$\mathrm{ms}[a_t^{\gamma^i}, \gamma^j] = \mathrm{ms}[a_t^{\gamma^\ell}, \gamma^\ell] + (\gamma^j - \gamma^\ell)^2 h_0 + (\gamma^i - \gamma^\ell)(2h_0^{\gamma^j} - 1) \ . \tag{2.9}$$

By substituting

$$\hat{n}_t^{\gamma^j} = \hat{n}_t^{\gamma^\ell} - \hat{e}_t^{\gamma^j,\gamma^\ell} = \hat{n}_t^{\gamma^\ell} - (\gamma^j - \gamma^\ell)\tilde{e}_t$$

into the defining expression $h_0^{\gamma^j} = \mathrm{cov}(\tilde{e}_t, \hat{n}_t^{\gamma^j})$, we can easily verify two companion formulas to (2.9),

$$h_0^{\gamma^j} = h_0^{\gamma^\ell} - (\gamma^j - \gamma^\ell)h_0 \qquad (2.10)$$

and

$$\partial \mathrm{ms}[a_t^{\gamma^i,\gamma^j}]/\partial\gamma^j = 2(\gamma^j - \gamma^i)h_0 \quad . \qquad (2.11)$$

Now we are prepared to solve the minimax problem (2.1). The formula (2.10) shows that $h_0^{\gamma^j}$ is a decreasing function of γ^j, so that there are three exhaustive and mutually exclusive possibilities: (i) $h_0^{\gamma^\ell} < 1/2$; (ii) $h_0^{\gamma^u} > 1/2$; and (iii) $h_0^{\gamma^u} \leqslant 1/2 \leqslant h_0^{\gamma^\ell}$. In case (i) we clearly have

$$\mathrm{ms}[a_t^{\gamma^i,\gamma^j}] < \mathrm{ms}[a_t^{\gamma^\ell,\gamma^j}] \quad , \qquad (2.12)$$

for all $\gamma^i > \gamma^\ell$, whereas in case (ii) the inequality

$$\mathrm{ms}[a_t^{\gamma^i,\gamma^j}] < \mathrm{ms}[a_t^{\gamma^u,\gamma^j}] \qquad (2.13)$$

holds for all $\gamma^i < \gamma^u$. For case (iii), we solve (2.10) for the unique γ^* for which $h_0^{\gamma^*} = 1/2$, obtaining

$$
\begin{aligned}
\gamma^* &= \gamma^\ell + h_0^{-1}(h_0^{\gamma^\ell} - 1/2) \\
&= \gamma + h_0^{-1}(h_0^{\gamma} - 1/2) \\
&= \gamma + (h_0^{\gamma'} - h_0^{\gamma})^{-1}(\gamma - \gamma')(h_0^{\gamma} - 1/2) \qquad (2.14)
\end{aligned}
$$

for any γ, $\gamma' \in [\gamma^\ell, \gamma^u]$ with $\gamma' \neq \gamma$. In case (iii) it is clear that (2.13) holds for $\gamma^j \in [\gamma^\ell, \gamma^*)$, that (2.12) holds for $\gamma^j \in (\gamma^*, \gamma^u]$, and that $ms[a_t^{\gamma^i, \gamma^*}]$ does not depend on γ^i. Finally, observe (for all cases) from (2.11) that $ms[a_t^{\gamma^i, \gamma^j}]$ for fixed γ^i, i.e., as a function of γ^j, is *uniquely* minimized over both $[\gamma^\ell, \gamma^i]$ and $[\gamma^i, \gamma^u]$ at their common endpoint γ^i. (One of these intervals will contain only a single number if γ^i is an endpoint of $[\gamma^\ell, \gamma^u]$.) These facts lead immediately to the following result.

Theorem. *The minimax problem* (2.1) *has a unique solution for* γ^j *which is given by* $\gamma^j = \gamma^\ell$ *if* $h_0^{\gamma^\ell} < 1/2$ (*minimum variance signal extraction*); *by* $\gamma^j = \gamma^u$ *if* $h_0^{\gamma^u} > 1/2$ (*maximum variance signal extraction*); *or by* $\gamma^j = \gamma^*$ *defined by* (2.14) *if* $h_0^{\gamma^u} < 1/2 < h_0^{\gamma^\ell}$.

The solution $\gamma^j = \gamma^*$ is particularly interesting, because of the fact that $ms[a_t^{\gamma, \gamma^*}]$ does not depend on γ. This means that a natural measure of seasonal adjustment standard error exists which is the same for all candidate choices n_t^γ, $\gamma \in [\gamma^\ell, \gamma^u]$ of the nonseasonal component.

For later use, we note that no special properties have been required of the endpoints γ^{ℓ} and γ^{u}. Further, if we replace γ^{ℓ} in (2.3) by any $\gamma' < \gamma$, we obtain a more general formula for the γ(and γ')-invariant random variable $\tilde{e}_t$, namely,

$$\tilde{e}_t = (\gamma - \gamma')^{-1}\hat{e}_t^{\gamma,\gamma'} . \qquad (2.15)$$

3. Obtaining $h_0^{\gamma} = h_0^{\gamma}(t)$ and $h_0 = h_0(t)$ from a Kalman Smoother

In a real data situation, only finitely many random variables $x_1,\ldots,x_T$ are observed. In this case, the coefficients $h_i^{\gamma} = h_i^{\gamma}(t)$ for which

$$n_t^{\gamma} = \sum_{i=-(T-t)}^{t-1} h_i^{\gamma} x_{t-i}$$

holds are not straightforward to calculate, especially when the x_t obey a nonstationary ARMA model. In this situation, it is natural to seek to represent the signal extraction problem in such a way that a (e.g., fixed-interval) Kalman smoother recursion algorithm (see Anderson and Moore (1979) can be applied to obtain $\hat{n}_t^{\gamma}$. In the next section we shall illustrate, by means of a thorough analysis of an elementary example, the kinds of problems which must be solved to obtain $\hat{n}_t^{\gamma}$ precisely, as opposed to approximately, via a Kalman smoother. Our concern now is to show with the aid of the formula (2.15) that the pivotal coefficient $h_0^{\gamma} = h_0^{\gamma}(t)$ can be obtained from the error covariance matrices produced by the algorithm, provided that for some $\gamma' < \gamma$, $e_t^{\gamma,\gamma'}$ is placed in the Kalman state vector along with n_t^γ, as in the example below. Indeed, the algorithm then produces $\mathrm{var}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'})$ and $\mathrm{cov}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'}, n_t^\gamma - \hat{n}_t^\gamma)$. The latter quantity is equal to $-\mathrm{cov}(\hat{e}_t^{\gamma,\gamma'}, \hat{n}_t^\gamma)$ because n_t^γ and $e_t^{\gamma,\gamma'}$ are uncorrelated and so are $n_t^\gamma - \hat{n}_t^\gamma$ and $\hat{e}_t^{\gamma,\gamma'}$. Thus we simply use (2.15) to obtain our desired result,

$$h_0^\gamma(t) = -(\gamma-\gamma')^{-1}\mathrm{cov}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'}, n_t^\gamma - \hat{n}_t^\gamma). \tag{3.1}$$

There is more, however, because we can determine $h_0(t) = \mathrm{var}(\tilde{e}_t)$ from $\mathrm{var}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'})$: Using (2.6) and the fact that $e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'}$ is uncorrelated with $\hat{e}_t^{\gamma,\gamma'}$, we have

$$\begin{aligned} h_0(t) &= (\gamma - \gamma')^{-2}\mathrm{var}(\hat{e}_t^{\gamma,\gamma'}) \\ &= (\gamma - \gamma')^{-2}\{\mathrm{var}(e_t^{\gamma,\gamma'}) - \mathrm{var}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'})\} \\ &= (\gamma - \gamma')^{-1} - (\gamma - \gamma')^{-2}\mathrm{var}(e_t^{\gamma,\gamma'} - \hat{e}_t^{\gamma,\gamma'}) \ . \end{aligned}$$

Because of the linear relation (2.10), we can use $h_0(t)$ and $h_0^\gamma(t)$ for a single value of γ to produce $h_0^\gamma(t)$ for all $\gamma \in [\gamma^\ell,\gamma^u]$. Thus, with the information from a run of a Kalman smoother for the decomposition $x_t = n_t^\gamma + e_t^{\gamma,\gamma'} + s_t^{\gamma'}$ associated with a <u>single</u> pair $\gamma,\gamma'(\gamma>\gamma')$, we can use the Theorem of section 2 to determine the solution of Watson's minimax problem for each $t=1,\ldots,T$!

4. An Example

Let B denote the backshift operator, $Bx_t = x_{t-1}$. Suppose we have twice-yearly observations $x_1, x_2, \ldots, x_T$ from a seasonal series conforming to the model

$$(I - B^2)x_t = \varepsilon_t \quad , \tag{4.1}$$

where ε_t is a white noise process with variance σ_ε^2. Suppose also that x_t is known to have the form $x_t = s_t + n_t$, where the seasonal component s_t and the nonseasonal component n_t are such that the transformed series $(I+B)s_t$ and $(I-B)n_t$ are stationary and uncorrelated. Different aspects of this example are discussed in Pierce and Maravall (1984). Using partial fractions, we obtain the pseudo-spectrum decompositions

$$f_x(\omega) = \frac{\sigma_\varepsilon^2}{|1 - e^{-i2\omega}|^2} = \frac{\sigma_\varepsilon^2/4}{|1 + e^{-i\omega}|^2} + \frac{\sigma_\varepsilon^2/4}{|1 - e^{-i\omega}|^2}$$

$$= \{ \frac{\sigma_\varepsilon^2}{4} |1 + e^{-i\omega}|^{-2} + \gamma\} + \{ \frac{\sigma_\varepsilon^2}{4} |1 - e^{-i\omega}|^{-2} - \gamma\}$$

$$= \qquad f_s^\gamma(\omega) \qquad + \qquad f_n^\gamma(\omega) \quad ,$$

for which we require only that the component pseudo-spectra $f_s^\gamma(\omega)$ and $f_n^\gamma(\omega)$ are non-negative. Both $|1 + e^{-i\omega}|^2 = 2(1 + \cos\omega)$ and $|1 - e^{-i\omega}|^2 = 2(1 - \cos\omega)$ have the maximum value 4, so this requirement means that $\gamma \in [-\sigma_\varepsilon^2/16, \sigma_\varepsilon^2/16]$.

For $\gamma \in (0,\sigma_\varepsilon^2/16]$, if we define $f_e^{\gamma,0}(\omega) = \gamma$, then

$$f_x(\omega) = f_s^0(\omega) + f_e^{\gamma,0}(\omega) + f_n^\gamma(\omega) \quad , \tag{4.2}$$

corresponding to the decomposition

$$x_t = s_t^0 + e_t^{\gamma,0} + n_t^\gamma \quad . \tag{4.3}$$

We shall formulate a Kalman state model for this decomposition. To do this we require the coefficients and innovation variances of ARMA models associated with the pseudo-spectra of the components. Corresponding to $f_s^0(\omega)$ we clearly have the model

$$(I + B)s_t^0 = \alpha_t^0 \tag{4.4}$$

where α_t^0 is a white noise process with variance $\sigma_0^2 = \sigma_\varepsilon^2/4$. Of course, $e_t^{\gamma,0}$ is a white noise process with variance γ. The situation is more complex for n_t^γ.

4.1 Spectral Factorization for n_t^γ .

Since $(I-B)n_t^\gamma$ has the spectrum

$$\begin{aligned} |1 - e^{-i\omega}|^2 f_n^\gamma(\omega) &= \sigma_\varepsilon^2/4 - \gamma|1 - e^{-i\omega}|^2 \\ &= \{\sigma_0^2 - 2\gamma\} + 2\gamma\cos\omega \end{aligned} \tag{4.5}$$

it is clearly an MA(1) process, i.e.,

$$(I-B)n_t^\gamma = (I - \theta^\gamma B)\beta_t^\gamma \tag{4.6}$$

where the coefficient θ^γ (satisfying $|\theta^\gamma|<1$) and the variance σ_γ^2 of the white noise process β_t^γ must be obtained from the spectral density function (4.5), a procedure known as spectral factorization. The procedure is easy for this example. From $\{\sigma_0^2 - 2\gamma\} + 2\gamma\cos\omega = \sigma_\gamma^2|1 - \theta^\gamma e^{-i\omega}|^2 = \sigma_\gamma^2(1 + \{\theta^\gamma\}^2) - 2\sigma_\gamma^2\theta^\gamma\cos\omega$, we obtain the system of equations

$$\sigma_0^2 - 2\gamma = \sigma_\gamma^2(1 + \{\theta^\gamma\}^2)$$

$$\gamma = -\theta^\gamma\sigma_\gamma^2 \ , \tag{4.7}$$

whose solution with $|\theta^\gamma|<1$ is given by

$$\theta^\gamma = -\left(\frac{\sigma_0^2}{2\gamma} - 1\right) + \left\{\left(\frac{\sigma_0^2}{2\gamma} - 1\right)^2 - 1\right\}^{1/2}$$

$$\sigma_\gamma^2 = (1 - \theta^\gamma)^{-2}\sigma_0^2 \ . \tag{4.8}$$

Note that $0 < \gamma \leq \sigma_0^2/4$ implies $-1 \leq \theta < 0$.

<u>Remark</u>. When x_t has a somewhat more complex model than (4.1), the spectral factorization required to determine the model for n_t^γ ($\gamma\neq 0$) could ordinarily only be approximated, using an algorithm like that of Wilson (1972), or a root-finding algorithm applied to the covariance generating function, see Whittle (1963, p. 28). (The latter procedure must be used when the

moving average component of the model is noninvertible. Noninvertibility occurs at the maximum admissible value γ^u of γ, since $f_n^0(\omega) - \gamma^u$ is zero for some ω.)

We can now put the modeling equations for x_t, s_t^0, n_t^γ and $e_t^{\gamma,0}$ in a "Markovian" form as required by the smoothing algorithms:

$$\begin{bmatrix} s_t^0 \\ n_t^\gamma \\ \beta_t^\gamma \\ e_t^{\gamma,0} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & -\theta^\gamma & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} s_{t-1}^0 \\ n_{t-1}^\gamma \\ \beta_{t-1}^\gamma \\ e_{t-1}^{\gamma,0} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_t^0 \\ \beta_t^\gamma \\ e_t^{\gamma,0} \end{bmatrix} . \qquad (4.9)$$

$$x_t = [\,1 \quad 1 \quad 0 \quad 1\,][s_t^0 \quad n_t^\gamma \quad \beta_t^\gamma \quad e_t^{\gamma,0}]' \ . \qquad (4.10)$$

We also require the covariance matrix of the "noise" vector $[\alpha_t^0\ \beta_t^\gamma\ e_t^{\gamma,0}]'$ in (4.9). If we multiply (4.2) by $|1 - e^{-i2\omega}|^2$, we obtain a decomposition of the spectral density of the white noise process $\varepsilon_t = z_t - z_{t-2}$,

$$\sigma_\varepsilon^2 = \sigma_0^2|1 - e^{-i\omega}|^2 + \gamma|1 - e^{-i2\omega}|^2$$
$$+ \sigma_\gamma^2|1 + e^{-i\omega}|^2|1 - \theta^\gamma e^{-i\omega}|^2 \quad ,$$

which reveals that the stationary processes $(I-B)\alpha_t^0$, $(I-B^2)e_t^{\gamma,0}$ and $(I+B)(I-\theta B)\beta_t^\gamma$ are uncorrelated. It can be shown to follow from this that the processes α_t^0, $e_t^{\gamma,0}$ and β_t^γ are uncorrelated, see Findley (1985) or

Whittle (1963, p. 44). Hence, in particular, the covariance matrix Q of $[\alpha_t^0\ \beta_t^\gamma\ e_t^{\gamma,0}]'$ is given by

$$Q = \begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & (1-\theta^\gamma)^{-2}\sigma_0^2 & 0 \\ 0 & 0 & \gamma \end{bmatrix} . \tag{4.11}$$

Now we must specify properties of the starting values s_1^0 and n_1^γ, without which the specifications of s_t^0 by (4.4), of n_t^γ by (4.6) and of x_t by (4.1) are incomplete. Bell (1984) discusses this point very informatively as well as the effect of the starting-value-covariance specifications on the form of the filters providing the least squares estimates of the components of x_t.

4.2 Choosing Covariance Properties for the Initial Values.

In order to obtain estimates $\hat{n}_t^\gamma$, $\hat{s}_t^0$, $\hat{e}_t^{\gamma,0}$ (and $\hat{\beta}_t^\gamma$) along with their associated error covariance matrix from a Kalman smoother, for t=1,...,T, we must have n_1^γ and s_1^γ uncorrelated with α_t^0, β_t^γ and $e_t^{\gamma,0}$ for t=2,...,T, as are β_1^γ and $e_1^{\gamma,0}$ by our previous discussion. In this case, the following requirement is met, i.e., the representation (4.9) is truly Markovian.

(I) <u>As regards its covariance structure</u>, $[s_t^0\ n_t^\gamma\ \beta_t^\gamma\ e_t^{\gamma,0}]'$ <u>is a Markov process for</u> t=2,...,T <u>initialized at</u> t=1.

This justifies the application of the Kalman smoothing algorithm for $t=1,\ldots,T$.

But two more requirements must be met by the processes satisfying (4.9), (4.10) (equivalently, (4.3-4) and (4.6)) if they are to have the properties of the decompositions described in section 2.

(II) The covariance structure of $x_1,\ldots,x_T$ is independent of the value of γ.

(III) $\operatorname{cov}(e_t^{\gamma,0}, n_u^{\gamma}) = \operatorname{cov}(e_t^{\gamma,0}, s_u^{0}) = 0,\ 1 \leq t,\ u \leq T.$
Also, $\operatorname{var}(n_t^{\gamma} + e_t^{\gamma,0}) = \operatorname{var}(n_t^{\gamma}) + \gamma$ does not depend on γ, for $t=1,\ldots,T$.

We will now show that when (I) holds, which we assume, then (II) and (III) are implied by the weaker conditions (II') and (III') concerning the initial values, which we shall impose.

(II') $\operatorname{var}(x_1)$, $\operatorname{var}(x_2)$ and $\operatorname{cov}(x_1,x_2)$ do not depend on γ.

(III') $\operatorname{cov}(s_1^0, e_1^{\gamma,0}) = \operatorname{cov}(n_1^{\gamma}, e_1^{\gamma,0}) = 0.$
Also, $\operatorname{var}(n_1^{\gamma} + e_1^{\gamma,0}) = \operatorname{var}(n_1^{\gamma}) + \gamma$ does not depend on γ.

In conformity with the notation of section 2, let us define $n_1^0 = n_1^{\gamma} + e_t^{\gamma,0}$. Since

$$\operatorname{var}(x_1) = \operatorname{var}(s_1^0) + \operatorname{var}(n_1^0) + 2\operatorname{cov}(s_1^0, n_1^0) \tag{4.12}$$

and since $\operatorname{var}(x_1)$, $\operatorname{var}(s_1^0)$ and $\operatorname{var}(n_1^0)$ do not depend on γ, there is a constant C such that

$$\mathrm{cov}(s_1^0,\ n_1^0) = \mathrm{cov}(s_1^0,\ n_1^\gamma) = C \quad . \tag{4.13}$$

Because $s_2^0 = -s_1^0 + \alpha_2^0$ by (4.4) and $n_2^\gamma = n_1^\gamma + \beta_2^\gamma - \theta^\gamma\beta_1^\gamma$ by (4.6), we have

$$x_2 = n_1^\gamma - s_1^0 + e_2^{\gamma,0} + \alpha_2^0 + \beta_2^\gamma - \theta^\gamma\beta_1^\gamma \ . \tag{4.14}$$

Hence

$$\begin{aligned}\mathrm{var}(x_2) &= \mathrm{var}(n_1^\gamma) + \mathrm{var}(s_1^0) - 2\ \mathrm{cov}(s_1^0,\ n_1^\gamma) + \gamma \\ &\qquad + \sigma_0^2 + (1 + \{\theta^\gamma\}^2)\sigma_\gamma^2 - 2\ \mathrm{cov}(n_1^\gamma - s_1^0,\ \theta^\gamma\beta_1^\gamma) \\ &= \mathrm{var}(x_1) - 4\ \mathrm{cov}(s_1^0,\ n_1^0) + 2\sigma_0^2 \\ &\qquad - 2\{\mathrm{cov}(n_1^\gamma - s_1^0,\ \theta^\gamma\beta_1^\gamma) + \gamma\} \quad ,\end{aligned} \tag{4.15}$$

by (I), (III'), (4.8) and (4.12).

Similarly

$$\begin{aligned}\mathrm{cov}(x_1,x_2) &= \mathrm{var}(n_1^0) - \mathrm{var}(s_1^0) \\ &\qquad - \{\mathrm{cov}(n_1^\gamma + s_1^0,\ \theta^\gamma\beta_1^\gamma) + \gamma\} \quad .\end{aligned} \tag{4.16}$$

The fact that $\mathrm{var}(x_1)$, $\mathrm{var}(x_2)$, $\mathrm{cov}(x_1,x_2)$, $\mathrm{var}(n_1^0)$, $\mathrm{var}(s_1^0)$ and $\mathrm{cov}(n_1^0,s_1^0)$ do not depend on γ means that the same must be true of the { }-terms in (4.15) and (4.16). Since these terms clearly approach 0 as γ (and hence also θ^γ, see (4.7)) approaches 0, they must, in fact, be equal to 0. Because $\gamma = -\theta^\gamma\sigma_\gamma^2$, this is equivalent to the formulas

$$\begin{aligned} \mathrm{cov}(s_1^0, \beta_1^\gamma) &= 0 \\ \mathrm{cov}(n_1^\gamma, \beta_1^\gamma) &= \sigma_\gamma^2 \end{aligned} \tag{4.17}$$

holding for all γ. Now, using (4.17), (4.8) and the model equations (4.1), (4.4), and (4.6), it is straightforward to verify that the white noise variable

$$\begin{aligned} \varepsilon_t &= x_t - x_{t-2} \\ &= \{e_t^{\gamma,0} - e_{t-2}^{\gamma,0}\} + \{\alpha_t^0 - \alpha_{t-1}^0\} \\ &\qquad + \{\beta_t^\gamma + (1 - \theta^\gamma)\beta_{t-1}^\gamma - \theta^\gamma\beta^\gamma{}_{t-2}\} \end{aligned} \tag{4.18}$$

is uncorrelated with x_1 and x_2 for t=3,...,T, and therefore that (II) holds. The verification of (III) also follows readily.

The conditions (III'), (4.13) and (4.17) reveal how the covariance matrix of the state vector $(s_1^0\ n_1^\gamma\ \beta_1^\gamma\ e_1^{\gamma,0})'$ initializing the Markov process (4.9) (and therefore the Kalman smoothing algorithm) must be restricted in order for (II) and (III) to be satisfied: The nonnegative value of $\mathrm{var}(s_1^0)$ can be selected arbitrarily. However, by (4.17) and the Cauchy-Schwarz inequality,

$$\begin{aligned} \sigma_\gamma^2 = \mathrm{cov}(n_1^\gamma, \beta_1^\gamma) &= \mathrm{cov}(n_1^\gamma + e_1^{\gamma,0}, \beta_1^\gamma) \\ &\leq \{\mathrm{var}(n_1^\gamma) + \gamma\}^{1/2}\sigma_\gamma \ . \end{aligned}$$

Since $\sigma_0^2 = \max_\gamma \sigma_\gamma^2$ by (4.7), the constant $\mathrm{var}(n_1^0) = \mathrm{var}(n_1^\gamma) + \gamma$ must be therefore chosen to be at least as large as σ_0^2. Similarly,

$$|\mathrm{cov}(n_1^\gamma, s_1^0)| = |\mathrm{cov}(n_1^0, s_1^0)| < \sqrt{\mathrm{var}(n_1^0)}\ \sqrt{\mathrm{var}(s_1^0)} \quad . \qquad (4.19)$$

This is the only restriction that must be imposed on the choice of $\mathrm{cov}(n_1^\gamma, s_1^0)$.

Sometimes, other considerations make additional restrictions appropriate. For example, in the model (4.1), the facts that x_{2m} depends only on $\varepsilon_{2m}, \varepsilon_{2m-2}, \ldots, \varepsilon_2$ and x_2, and that x_{2m-1} depends only on $\varepsilon_{2m-1}, \varepsilon_{2m-3}, \ldots, \varepsilon_1$ and x_1 ($m \geq 2$) makes it seem natural to choose the covariance structure of n_1^γ and s_1^0 in such a way that $\mathrm{cov}(x_1, x_2) = 0$ and $\mathrm{var}(x_1) = \mathrm{var}(x_2)$. By (4.15) - (4.17), this means choosing $\mathrm{var}(n_1^0) = \mathrm{var}(s_1^0)$ and $\mathrm{cov}(n_1^\gamma, s_1^0) = \sigma_0^2/2$. Thus the covariance matrix of $(s_1^0\ n_1^\gamma\ \beta_1^\gamma\ e_1^{\gamma,0})'$ is required to be

$$P_0^\gamma = \begin{bmatrix} \mathrm{var}(s_1^0) & \sigma_0^2/2 & 0 & 0 \\ \sigma_0^2/2 & \mathrm{var}(s_1^0)-\gamma & (1-\theta^\gamma)^{-2}\sigma_0^2 & 0 \\ 0 & (1-\theta^\gamma)^{-2}\sigma_0^2 & (1-\theta^\gamma)^{-2}\sigma_0^2 & 0 \\ 0 & 0 & 0 & \gamma \end{bmatrix}$$

We must choose $\mathrm{var}(s_1^0) \geq \sigma_0^2$, and then (4.19) is also satisfied. Often, setting $\mathrm{var}(s_1^0)$ equal to some fraction of the squared range $(\max_t\{x_t\} - \min_t\{x_t\})^2$ would provide a conservative choice.

It is clear that the analysis which led us to P_0^γ could be very difficult to carry out for a more complicated model than (4.1). If the backshift-operator polynomial transformations to stationarity for s_t^γ and n_t^γ ((I+B) and (I-B) in our example) have no common factor, then the estimation error covariance matrices will converge as $t\to\infty$ to a matrix which is independent of the initializing covariance matrix, see Burridge and Wallis (1983), for example. Thus, the same will be true of $h_0^\gamma(t)$, by (3.1). In the seasonal adjustment situation, the greatest interest would usually be attached to h_t^γ for values of t close to T. In this case, it therefore seems reasonable, if T is not too small, to use a simpler procedure for obtaining an initializing covariance matrix.

We will briefly illustrate one such procedure using our decomposition $x_t = s_t^0 + e_t^{\gamma,0} + n_t^0$ of the series satisfying (4.1). Following Assumption A of Bell(1984), we will choose initial values x_1 and x_2 which are uncorrelated with the white noise series α_t^0, β_t and $e_t^{\gamma,0}$ for all t (contradicting (III), which implies via (4.17) that $\mathrm{cov}(x_2,\beta_1) = \gamma$) and whose variance-covariance matrix is specified independently of γ. Then we solve (4.3) and (4.14) for the initial values s_1^0 and n_1^γ, obtaining

$$s_1^0 = \{x_1 - x_2 + \alpha_2^0 + \beta_2^\gamma - \theta^\gamma\beta_1^\gamma - e_1^{\gamma,0} + e_2^{\gamma,0}\}/2$$

$$n_1^\gamma = \{x_1 + x_2 - \alpha_2^0 - \beta_2^\gamma + \theta^\gamma\beta_1^\gamma - e_1^{\gamma,0} - e_2^{\gamma,0}\}/2 \; . \quad (4.20)$$

The variance-covariance matrix of s_1^0 and n_1^γ, and therefore also of s_t^0 and n_t^γ for $t\geq 2$, can be calculated from the known covariances of the right-hand-side terms in (4.20). It is easy to see that (II) holds, and if we initialize (4.9) at $t=2$ instead of $t=1$, so does the appropriately modified version of (I). Thus a Kalman smoother can provide least mean square estimates of n_t^γ, s_t^0 etc. for $t\geq 2$ based on the shortened observation set $\{x_2,\ldots,x_T\}$. Finally, it is simple to verify that all of the conditions of (III) are satisfied for $t\geq 3$, but not for $t=1$ or 2.

5. Watson's Second Minimax Criterion.

We return to the notation and discussion of Section 2. Since the least mean square approximation in OBS to $\Delta n_t^\gamma = n_t^\gamma - n_{t-1}^\gamma$ is given by $\Delta\hat{n}_t^\gamma = \hat{n}_t^\gamma - \hat{n}_{t-1}^\gamma$ (which is different, in general, from the orthogonal projection of Δn_t^γ onto the linear subspace spanned by Δx_{t-i}, $-m+1 \leq i \leq n$), it makes sense to consider the criterion

$$\min_{\gamma^j} \max_{\gamma^i} \mathrm{ms}[\Delta a_t^{\gamma^i,\gamma^j}] \tag{5.1}$$

if the focus is on estimation of Δn_t. Watson's results on (5.1) for a special case can be obtained in the present generality using the analogue of (2.9) given by (5.3) below. First, note that by arguments similar to those used to derive (2.7), the identity

$$\Delta n_t^{\gamma^\ell} - \Delta\hat{y}_t = \Delta n_t^{\gamma^i} - \Delta\hat{y}_t + \Delta e_t^{\gamma^i,\gamma^\ell}$$

leads to

$$E\{\Delta n_t^{\gamma^\ell} - \Delta\hat{y}_t\}^2 = E\{\Delta n_t^{\gamma^i} - \Delta\hat{y}_t\}^2$$
$$+ 2(\gamma^i - \gamma^\ell)\{1 - \mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{y}_t)\} \quad . \qquad (5.2)$$

By (2.5) the quantity $\mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{y}_t)$ is equal to $v_0(t) + v_0(t-1) - v_{-1}(t-1) - v_1(t)$, when $\hat{y}_t$ is given by (2.4) with coefficients $v_i = v_i(t)$. Then, proceeding as before with $\hat{y}_t = \hat{n}_t^{\gamma^j}$ and setting $h_1(t) = \mathrm{cov}(\tilde{e}_t, \tilde{e}_{t-1})$, one readily obtains

$$\mathrm{ms}[\Delta a_t^{\gamma^i}, \gamma^j] = \mathrm{ms}[\Delta a_t^{\gamma^\ell}, \gamma^\ell] + \{h_0(t) + h_0(t-1) - 2h_1(t)\}(\gamma^j - \gamma^\ell)^2$$
$$+ 2(\gamma^i - \gamma^\ell)\{\mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{n}_t^{\gamma^j}) - 1\} \quad , \qquad (5.3)$$

with $h_0(t) + h_0(t-1) > 2h_1(t)$ unless $\tilde{e}_t = \tilde{e}_{t-1}$. However, $\tilde{e}_t = \tilde{e}_{t-1}$ is equivalent to $\hat{n}_t^\gamma - \hat{n}_t^{\gamma'} = \hat{n}_{t-1}^\gamma - \hat{n}_{t-1}^{\gamma'}$ for all $\gamma, \gamma' \in [\gamma^\ell, \gamma^u]$ with $\gamma > \gamma'$, and so can, in principle, be verified or contradicted from calculated quantities. Since

$$\mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{n}_t^{\gamma^j}) = \mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{n}_t^{\gamma^\ell})$$
$$- \{h_0(t) + h_0(t-1) - 2h_1(t)\}(\gamma^j - \gamma^\ell) \ ,$$

the form of (5.3) is equivalent to that of (2.9) for analyzing the mimimax criterion and the analogues of the conclusions of the Theorem of section 2 therefore hold for (5.1) when $\tilde{e}_t \neq \tilde{e}_{t-1}$. Also, a formula like (2.14) can be given for the value γ^{**} of γ^j for which the coefficient of $(\gamma^i - \gamma^\ell)$ in (5.3) vanishes.

In analogy with the derivations of section 3, the quantities $h_1(t)$ and $\mathrm{cov}(\Delta\tilde{e}_t, \Delta\hat{n}_t^\gamma)$ can be obtained for all $t \geq 2$ from the Kalman smoother, provided that the state vector at time t contains $e_t^{\gamma,\gamma'}$, $e_{t-1}^{\gamma,\gamma'}$ and n_{t-1}^γ among its components, along with n_t^γ. The effect of the presence of these additional components on the initializing covariance matrix is easy to determine for the example of section 4. To maintain compatibility with the equations (4.4) and (4.6) and the conditions (II) and (III), one should specify the covariance matrix of $[e_0^{\gamma,0}\ n_0^\gamma]$ with the other initial components $[s_1^0\ n_1^\gamma\ \beta_1^\gamma\ e_1^{\gamma,0}]$ to be

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ \mathrm{cov}(s_1^0, n_1^0) & \mathrm{var}(n_0^\gamma) + \gamma & 0 & 0 \end{bmatrix},$$

and also specify $\mathrm{var}(e_0^{\gamma,0}) = \gamma$, and

$$\mathrm{var}(n_0^\gamma) = \mathrm{var}(n_1^\gamma) - \sigma_\gamma^2(1 - \theta^\gamma)^2 \quad .$$

If Bell's Assumption A is to be used, then the covariance of n_0^γ with the other components can be determined with the aid of (4.20) and the equation

$$n_0^\gamma = \{x_2 + x_1 + \alpha_1^0 - \alpha_2^0 - \beta_2^\gamma$$
$$- (2 - \theta^\gamma)\beta_1^\gamma + 2\theta^\gamma\beta_0^\gamma + e_2^{\gamma,0} - e_1^{\gamma,0}\}/2 \quad ,$$

which is obtained from an analogue of (4.20) and the relation $x_0 = x_2 - \varepsilon_2$, using (4.18) with t=2.

Concluding Remarks.

In this note we have demonstrated the great generality of the minimax results of Watson (1984) for selecting an unobserved components decomposition of an observed series when there is ambiguity concerning the assignment of a white noise process which is uncorrelated with the remaining components. We have also shown how the minimax choice can be determined from the output of a Kalman smoothing algorithm when (possibly nonstationary) ARMA models are available for the components of a candidate decomposition, provided that the initializing covariance matrix is properly chosen. Regardless of how the component estimates are calculated, the results of Bell (1984) show that if the component series are nonstationary, then initializing covariance assumptions must be made, explicitly or implicitly, before the minimum mean square estimates can be determined. The example of section 4 suggests that it will be difficult, in general, to determine which initializing assumptions are completely compatible with the assumptions (II) and (III) of section 4 concerning the component decomposition, but in practice mild incompatibilities may be harmless. Further work is needed to establish practical, adequate procedures for specifying initial covariances.

Acknowledgements

The author is grateful to William Bell, Peter Burman and Mark Watson for comments on a draft of this paper. The views expressed are those of the author and do not necessarily reflect those of the Census Bureau.

References

Anderson, Brian D.O. and John B. Moore (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.

Burridge, Peter and Kenneth F. Wallis (1983). Signal Extraction in Nonstationary Series. Warwick Economic Research Paper No. 234.

Findley, David F. (1985). On Backshift-Operator Polynomial Transformations to Stationarity for Nonstationary Series and Their Aggregates. Communications in Statistics, A(14), No. 1.

Hillmer, Steven C., William Bell and George C. Tiao (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 the Census.

Pierce, David A. and Augustin Maravall (1984). Advances in Model-Based Seasonal Adjustment. To appear in Proceedings of the Economic and Business Statistics Section, Washington, D.C.: American Statistical Association.

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

Whittle, Peter (1963). Prediction and Regulation. Princeton: Van Nostrand.

Wilson, G. Tunnicliffe (1969). Factorisation of the Generating Function of a Pure Moving Average Process. SIAM Journ. Num. Analysis. 6, 1-7.