Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 50 (2006) 2206 – 2231
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

An iterated parametric approach to nonstationary signal extraction

Tucker McElroy[a,*], Andrew Sutcliffe[b]

[a]*Statistical Research Division, US Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-9100, USA*
[b]*Australian Bureau of Statistics, Australia*

Available online 1 August 2005

Abstract

Consider the three-component time series model that decomposes observed data (Y) into the sum of seasonal (S), trend (T), and irregular (I) portions. Assuming that S and T are nonstationary and that I is stationary, it is demonstrated that widely used Wiener–Kolmogorov signal extraction estimates of S and T can be obtained through an iteration scheme applied to optimal estimates derived from reduced two-component models for S plus I and T plus I. This "bootstrapping" signal extraction methodology is reminiscent of the iterated nonparametric approach of the US Census Bureau's X-11 program. The analysis of the iteration scheme provides insight into the algebraic relationship between full model and reduced model signal extraction estimates.
Published by Elsevier B.V.

Keywords: ARIMA component model; Nonstationary time series; Seasonal adjustment; Signal extraction; Wiener–Kolmogorov filtering; X-11

1. Introduction

Extraction of a nonstationary signal from an observed, finite sample time series is a problem of both theoretical and practical interest. Work on stationary signal extraction from a signal plus noise model for an infinite sample dates back to Wiener (1949) and Kolmogorov (1939, 1941), whose celebrated solution has become classical in the time series literature. However, in many realistic situations, such as the project of deseasonalizing

* Corresponding author. Tel.: +1 301 763 3227; fax: +1 301 457 2299.
E-mail address: tucker.s.mcelroy@census.gov (T. McElroy).

0167-9473/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.csda.2005.07.008

economic data, the ambient signal is a nonstationary stochastic process. Essentially the same Wiener–Kolmogorov filter gives optimal extractions when the signal can be made stationary by appropriate differencing, and the noise is stationary—Cleveland and Tiao (1976) obtained results for this case. Also see Hannan (1967) and Sobel (1967) for earlier work. However, when the noise process is itself nonstationary—for example, in seasonal component estimation the noise consists of trend plus irregular—the situation becomes more complicated. Bell (1984) brought this issue to the forefront with an important paper, which produced extraction estimates under a variety of assumptions; in particular, Bell demonstrated that in order to obtain optimal estimates (in the sense of mean-squared error), it was essential to make assumptions about the data-generation process, such as Assumption A.

Assumption A states that the initial observed values are probabilistically independent of the differenced signal and differenced noise processes. When the differencing operators for the signal and noise components have no unit roots in common, Assumption A has the consequence that the initial values are independent of forecasts made from the differenced data—an assumption that is often made in the modelling and forecasting of time series. Assumption A is the approach that is implicitly adopted in the literature. One of the reasons is that given above—namely, future values of the differenced series are independent of the initial observed data; a second reason is that the formulas for the optimal signal extraction estimates are much simpler analytically—Bell (1984) shows that these formulas are analogous to those used in the stationary components scenario. A third appeal of Assumption A is that we are not required to know the covariance matrix of the initial conditions of the nonstationary process; fourthly, signal extraction estimates obtained under Assumption A are also locally optimal when Assumption A is removed—namely, those signal extractions are optimal (in the sense of having minimal mean squared error) within the class of linear functions of the data such that the error in the estimate does not depend on the initial values. This property underlies the "transformation approach" of Ansley and Kohn (1985), and is appealing because no assumptions need be made on the data-generation process. See also Kohn and Ansley (1986, 1987), Bell and Hilmer (1991), De Jong (1991), Koopman (1997), and Durbin and Koopman (2001) for implementations of the Kalman filter and smoother to produce estimates that are optimal under Assumption A.

In a three component model—consisting of trend, seasonal, and irregular portions—used to describe economic data, quite often the trend and seasonal are modelled as nonstationary processes, whereas the irregular is stationary. If one is interested in obtaining the trend, one must use signal extraction methods for a nonstationary signal (the trend) plus a nonstationary noise (the seasonal plus irregular) component. Under Assumption A, the finite-sample matrix formulas of Bell and Hilmer (1988) and Bell (2004) can be used; equivalently, a state space smoother (see Anderson and Moore, 1979) can produce the trend estimate once a model has been specified for each component. Of course, other components like a cycle can be handled by the Kalman smoother.

Another approach is to first detrend the data, by using trend extraction methods for a reduced "trend plus irregular" model; although this is an inaccurate depiction of reality, the matrix form of this estimate is easy to write down, since the noise (i.e., the irregular) is now stationary. After subtracting off this pilot trend estimate, one can then extract the seasonal component using a reduced "seasonal plus irregular" model—again, this model is not true to reality. However, one may iterate this algorithm and hope for convergence to the optimal signal extraction estimates. The Census Bureau program X-11 follows a similar procedure under a nonparametric umbrella, but one could conceive of implementing an analogous algorithm with parametric models for each of the components. This paper explores such an algorithm, and shows that iterations of these reduced model filters converge rapidly to the signal extraction filters that are optimal linear estimates under Assumption A.

This work is appealing on several grounds. It provides a natural, intuitive approach to the construction of optimal signal extraction estimates, built up from the less complicated filters coming from the stationary noise theory. In particular, the complicated initial value estimates of the trend and seasonal signals are automatically produced by the iterative structure of the algorithm presented in Section 4. This means that very simple computer programs can compute signal extraction estimates that are identical to those produced by the Kalman filter. Our procedure has the advantage of being easy to explain (no knowledge of State Space Models is required) and simple to implement. From a theoretical perspective, this paper's results provide insight into the algebraic relationship between trend and seasonal extraction.

Besides these aesthetic insights into signal extraction, analysis of this paper's main algorithm reveals the rate of convergence of iterative methods. Many practitioners in the economics and engineering communities will apply certain bandpass or lowpass filters to extract seasonal (or cycle) and trend components respectively. The most popular in the econometrics community are the Butterworth and Hodrick–Prescott filters—see Hodrick and Prescott (1997), which are essentially designed for stationary noise models, i.e., they are minimum mean squared error optimal for certain models that have nonstationary signal and stationary noise. Also see Pollock (2000, 2001) for examples of filtering nonstationary time series. Typically an economic practitioner interested in cycle estimation will apply a low-pass filter, remove the estimated trend, and then follow up with a band-pass filter to estimate the cycle. But this procedure does *not* produce the optimal estimate of the cycle, because in the first step a trend plus irregular model is assumed in lieu of the actual three component model. See Harvey and Trimbur (2003) for a discussion of these points. In fact, this consecutive use of low-pass and band-pass filters is identical to the first iteration of the general algorithm explored in this paper—an algorithm that converges exponentially fast. Thus, analysis of the convergence of our algorithm provides insight into how close the above common practice takes one to optimality. In the conclusion, we also compare this paper's algorithm to the X-11 procedure.

This paper first sets up nonstationary signal extraction for a three component model, giving full matrix formulas for various estimation procedures. This material, although mostly obtained from Bell and Hilmer (1988), is somewhat new in its formulation. Section 2 develops these formulas and the attendant notations and gives a motivating example. Section 3 discusses the main original theoretical results—namely the mathematical relationship between full model and reduced model signal extraction matrices. The iterative algorithm, which builds up the optimal full model filters from the reduced model filters, is analyzed in Section 4, and its rate of convergence is assessed through matrix norms. Section 5 discusses the implementation of these ideas, and presents the results of a simulation study and the analysis of Retail Sales of US Shoe Stores data. We conclude in Section 6, and provide one technical proof in the appendix.

2. Background and notation

We follow the notation of Bell and Hilmer (1988), and all formulas are presented in a vector framework. Thus, a sample of n observed data $Y_1, Y_2, \ldots, Y_n$ will be denoted by the column vector Y. Our basic model is

$$Y = S + T + I,$$

where Y is observed data, S represents the seasonal component (this should not be confused with the common use of S for signal), T is the trend, and I is the irregular. If we use the notation X_t for some stochastic process $\{X_t\}$, then we denote a single variate; if we just write X then we refer to the whole finite sample of $\{X_t\}$ written as a vector. We make the following assumptions:

- All covariance matrices are assumed to be invertible.
- The differenced seasonal, the differenced trend, and the irregular series are uncorrelated with one another. This will be referred to as the orthogonality property of the components.
- Both S and T are nonstationary, with associated differencing polynomials δ^S and δ^T, respectively that have distinct roots. Their orders are d_S and d_T, respectively, and we let $d = d_S + d_T$.
- The irregular component I is stationary.
- The data have a multivariate normal distribution.

A separate assumption, which we will sometimes impose below, is essentially Assumption A of Bell (1984), applied to a three component model: we assume that the initial values $Y_1, \ldots, Y_d$ are independent of the differenced trend, differenced seasonal, and the irregular.

Because the roots of δ^S and δ^T are distinct, their product is the minimal degree d polynomial δ, which is sufficient to reduce Y to stationarity, i.e.,

$$\delta(B) Y_t = W_t$$

is a stationary stochastic process (B denotes the backshift operator). The integration power series are simply the algebraic inverses of the differencing polynomials:

$$\xi(z) = 1/\delta(z) \quad \xi^T(z) = 1/\delta^T(z) \quad \xi^S(z) = 1/\delta^S(z).$$

If we wish to extract the seasonal, then we would write

$$W_t = \delta^T(B) U_t^S + \delta^S(B) V_t^S,$$

where $U_t^S = \delta^S(B) S_t$ is the differenced seasonal (in this case our signal), and $V_t^S = \delta^T(B)(T_t + I_t)$ is our differenced noise. Notice that the superscripts on U and V are necessary to distinguish differenced signal and noise for the seasonal extraction problem and the trend extraction problem, i.e., if we wish to extract trend, then we decompose as

$$W_t = \delta^S(B) U_t^T + \delta^T(B) V_t^T$$

where $U_t^T = \delta^T(B) T_t$ is our differenced signal and $V_t^T = \delta^S(B)(S_t + I_t)$ is our differenced noise.

Next, we develop filters from two-component models—either trend plus irregular or seasonal plus irregular. These models will be called *reduced models*, and typically do not represent reality. For example, the additive X-11 procedure initially assumes a trend plus irregular model for the data—even though this is unrealistic—in order to obtain initial trend estimates. For notation, write

$$
\begin{aligned}
Y_t^T &= T_t + I_t, \\
Y_t^S &= S_t + I_t
\end{aligned}
$$

for the two reduced models. In the first, δ^T is the appropriate differencing operator, while δ^S is appropriate for the second model. Note that signal extraction is much simpler for these reduced models, since estimation of a nonstationary signal (either T or S in these scenarios) when *stationary* noise (i.e., the irregular I) is present is reasonably straightforward. In particular, there is no explicit estimation of initial values of the nonstationary signal (this is performed implicitly through estimation of the stationary noise's initial values) as described in Bell and Hilmer (1988). If we apply the differencing operators to the reduced models, we obtain:

$$
\begin{aligned}
W_t^T &= \delta^T(B)Y_t^T = U_t^T + \delta^T(B)I_t, \\
W_t^S &= \delta^S(B)Y_t^S = U_t^S + \delta^S(B)I_t.
\end{aligned}
$$

We seek a relationship between the trend and seasonal extraction filters for the reduced models and the analogous filters for the full model. For practicality, all relationships are explored in a matrix form, since this is appropriate for finite samples. The estimates that we will consider are motivated by the following:

$$
\begin{aligned}
\mathbb{E}\left(T|Y^T\right) &= F_{TI}^T Y^T, \\
\mathbb{E}\left(S|Y^S\right) &= F_{SI}^S Y^S.
\end{aligned}
\tag{1}
$$

Because of the normality assumption on the stochastic process, these conditional expectations are given by linear operators acting on the data. If normality fails, one can still use the linear estimates, but there is no guarantee that they yield the conditional expectation. In other words, the conditional expectations of the signals T and S under their respective reduced models are given by the left multiplication of certain n by n filter matrices acting on data vectors Y^T and Y^S, respectively. Note that the expectation operators are different, because each reduced model will employ different assumptions; however, by an abuse of notation we will just use one operator $\mathbb{E}$. This point is further discussed below. We use the letter F for "filter," with superscript denoting the desired signal and subscript referencing the model—*SI* for Y^S, *TI* for Y^T, and *STI* for Y. Hence, our full model estimates are defined by

$$
\begin{aligned}
\mathbb{E}(T|Y) &= F_{STI}^T Y, \\
\mathbb{E}(S|Y) &= F_{STI}^S Y.
\end{aligned}
$$

The main result of this paper is to produce an elegant mathematical relationship between these various matrices F; associated with these formulas is a simple algorithm that will build up nonstationary signal extraction estimates for the full model completely from the overly simplistic reduced model estimates.

In order to express these matrices F explicitly, we must use appropriate matrix versions of the differencing operators. Following the approach of Pollock (2002), let L be a lag matrix—the finite-sample analog of the lag operator—defined by

$$L_{ij} = \begin{cases} 1 & \text{if } i = j+1, \\ 0 & \text{else.} \end{cases}$$

Then the Toeplitz differencing matrices are $\overline{\Delta}_T = \delta^T(L)$, $\overline{\Delta}_S = \delta^S(L)$, and $\overline{\Delta} = \delta(L)$. We define the "primary partition" to be

$$\overline{\Delta}_T = \begin{bmatrix} Q_T & 0 \\ & \Delta_T^* \end{bmatrix},$$

where Δ_T^* is a $n-d_T$ by n matrix, and Q_T is square d_T dimensional and lower triangular. We make similar definitions for $\overline{\Delta}$ and $\overline{\Delta}_S$. The Δ_T^* matrix operates on the data Y and produces trend-differenced data—hence the loss of dimension. But we must also consider applying trend differencing to data that has already been seasonally differenced; this will take $n-d_S$ dimensional data vectors and produce $n-d$ dimensional data that have been both seasonally and trend differenced. Accordingly, we define our "secondary partition" to be

$$\overline{\Delta}_T = \begin{bmatrix} R_T & P_T \\ 0 & \Delta_T \end{bmatrix},$$

where Δ_T is a $n-d$ by $n-d_S$ matrix, R_T is d_S by d dimensional, and P_T is d by $n-d_S$ dimensional. From the definitions of these two partitions, we see that

$$\begin{aligned} &U^T = \Delta_T^* T, \quad V^T = \Delta_S^*(S+I), \\ &U^S = \Delta_S^* S, \quad V^S = \Delta_T^*(T+I), \\ &W^T = \Delta_T^* Y^T, \quad W^S = \Delta_S^* Y^S, \\ &W = \Delta_S U^T + \Delta_T V^T = \Delta_T U^S + \Delta_S V^S. \end{aligned}$$

Now it follows from the commutativity of polynomial multiplication that

$$\overline{\Delta}_T \overline{\Delta}_S = \overline{\Delta} = \overline{\Delta}_S \overline{\Delta}_T. \tag{2}$$

The following lemma establishes a similar property for the differencing matrices that emerge from the two partitions:

Lemma 1.

$$\Delta_T \Delta_S^* = \Delta = \Delta_S \Delta_T^*.$$

Proof. Let $J_{0,d-1}$ be the selection matrix that picks off the last $n-d$ rows of an n by n matrix—so $J_{0,d-1}=\left[0_{n-d\times d}\ \ 1_{n-d\times n-d}\right]$. Then, using the secondary partition and applying the selection matrix to (2), we obtain

$$\begin{aligned}&[0\ \ \varDelta_T]\left[Q_S'\varDelta_S^{*\prime}\right]' = [0\ \ \varDelta] = [0\ \ \varDelta_S]\left[Q_T'\varDelta_T^{*\prime}\right]',\\ &\left[0\ \ \varDelta_T\varDelta_S^*\right] = [0\ \ \varDelta] = \left[0\ \ \varDelta_S\varDelta_T^*\right]\end{aligned}$$

which completes the proof. □

This provides us with new expressions for W:

$$\begin{aligned}W &= \varDelta_T U^S + \varDelta_S W^T\\ &= \varDelta_S U^T + \varDelta_T W^S\\ &= \varDelta_T U^S + \varDelta_S U^T + \varDelta I.\end{aligned}$$

It will be convenient to observe that $V^T = W^S$ and $V^S = W^T$, which can be checked directly.

Example. We flesh out these notations through an airline model example. The airline model is given by

$$(1-B)^2 U(B) Y_t = (1-\theta B)\left(1-\Theta B^{12}\right)\varepsilon_t^Y \tag{3}$$

for monthly data Y_t. The polynomial $U(z)=\left(1-z^{12}\right)/(1-z)$ is the seasonal summation operator. The model parameters are θ, Θ, and the variance of the white noise sequence ε_t^Y. So our various differencing operators are:

$$\begin{aligned}&\delta^T(B) = (1-B)^2, \quad \delta^S(B) = U(B) = 1 + B + \cdots + B^{11},\\ &\delta(B) = (1-B)\left(1-B^{12}\right).\end{aligned}$$

Note that these have distinct roots, which is important for the results of Bell and Hilmer (1988). It follows from (3) that

$$W_t = (1-\theta B)\left(1-\Theta B^{12}\right)\varepsilon_t^Y$$

and the inverses of the differencing polynomials (which may be power series) can be computed, e.g.,

$$\xi^T(z) = (1-z)^{-2} = 1 + 2z + 3z^2 + \cdots \ .$$

In practice, a modeler needs to decompose (3) into its component parts. If we form the canonical decomposition, assuming admissibility (see Hilmer and Tiao, 1982), then we obtain component models

$$\begin{aligned}&U(B)S_t = \theta^S(B)\varepsilon_t^S = U_t^S,\\ &(1-B)^2 T_t = \theta^T(B)\varepsilon_t^T = U_t^T,\\ &I_t = \varepsilon_t^I\end{aligned}$$

for various independent white noise sequences ε_t^S, ε_t^T, and ε_t^I. We have made identifications with differenced signals U_t in the last equalities. Finally, the differenced noises are

$$V_t^S = (1-B)^2\,(T_t + I_t) = \theta^T(B)\varepsilon_t^T + (1-B)^2 I_t,$$
$$V_t^T = U(B)\,(S_t + I_t) = \theta^S(B)\varepsilon_t^S + U(B) I_t.$$

All of these equations can be put into matrix form by replacing the operator B with the lag matrix L, and replacing the time series by column vectors. Computer software exists to estimate the airline model parameters and compute the component model parameters—the authors used the Ox programming language, along with various SsfPack routines (Doornik, 1998; Koopman et al., 1999).

Our next task will be to give explicit formulas for the F matrices. Bell and Hilmer (1988) provide expressions for the estimates, but here we write down the matrices explicitly. Also see De Jong (1991), Koopman (1997), and Durbin and Koopman (2001) for work that is similar in spirit. In the sequel we let 1_n denote the n by n identity matrix, and generally Σ_X denotes the covariance matrix for a random vector X. For any matrix M, we denote its transpose by M'. Before proceeding, we observe a complex issue pointed out to us by Bill Bell. Whereas we define the reduced model filter extraction matrices by applying Assumption A to both reduced models (and that this leads to the correct filter relationships is borne out by Theorem 1), these assumptions are not compatible with Assumption A on the full model. Hence, the use of Assumption A on the reduced models in the following proposition should be seen purely as a motivation for the derivation of the appropriate reduced model filters.

Proposition 1. *If we make Assumption A for the reduced models, we can write down a simple formula for either F_{TI}^T or F_{SI}^S. These are given by*

$$F_{TI}^T = 1_n - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* = \left(1_n + \Sigma_I \Delta_T^{*\prime} \Sigma_{U^T}^{-1} \Delta_T^*\right)^{-1},$$
$$F_{SI}^S = 1_n - \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* = \left(1_n + \Sigma_I \Delta_S^{*\prime} \Sigma_{U^S}^{-1} \Delta_S^*\right)^{-1}. \tag{4}$$

Proof. From Eq. (4.4) of Bell and Hilmer (1988), we have

$$F_{TI}^T Y = E\left(T|Y^T\right) = Y - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} W^T = \left(1_n - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^*\right) Y.$$

For the second equality of (4), which shows that F_{TI}^T is invertible, use the Sherman–Morrison –Woodbury formula (see Golub and Van Loan, 1996)

$$(A + UV')^{-1} = A^{-1} - A^{-1} U \left(1 + V' A^{-1} U\right)^{-1} V' A^{-1}$$

on the formula

$$\Sigma_{W^T} = \Sigma_{U^T} + \Delta_T^* \Sigma_I \Delta_T^{*\prime}$$

to obtain (with $A = \Sigma_{U^T}$, $U = \Delta_T^* \Sigma_I$, and $V' = \Delta_T^{*\prime}$)

$$\Sigma_{W^T}^{-1} = \Sigma_{U^T}^{-1} - \Sigma_{U^T}^{-1} \Delta_T^* \Sigma_I \left(1_n + \Delta_T^{*\prime} \Sigma_{U^T}^{-1} \Delta_T^* \Sigma_I\right)^{-1} \Delta_T^{*\prime} \Sigma_{U^T}^{-1}.$$

Then apply to this the matrix identity

$$1 - AB + ABA(1+BA)^{-1}B = (1+AB)^{-1}$$

which holds if A or B is invertible. This establishes the invertibility of F_{TI}^{T} and explicitly provides the inverse. A similar proof yields the expressions for F_{SI}^{S} as well. □

Remark 1. We have essentially used Assumption A only to derive formulas for the filters we wish to use. It follows from the "transformation approach" that the filters given by (4) are optimal signal extraction filters within the class of linear estimators whose error does not depend on the initial values, even when Assumption A does not hold. However, if Assumption A is true as well, then these filters (4) are also globally optimal.

In order to express the full model extraction matrices, we utilize formulas originally developed in Bell and Hilmer (1988) and reproduced in Bell (2004), which are repeated here for convenient reference. Then the trend extraction estimate, under Assumption A, is given by

$$\mathbb{E}(T|Y) = \overline{\Delta}_T^{-1} \begin{bmatrix} Q_T \hat{T}_* \\ \hat{U^T} \end{bmatrix},$$

where $\hat{T}_*$ is a d_T dimensional vector that estimates the initial values of the trend, and $\hat{U^T}$ is an estimate of the differenced trend derived from the differenced data W, i.e., according to (4.7) of Bell and Hilmer (1988),

$$\hat{U^T} = \Sigma_{U^T} \Delta_S' \Sigma_W^{-1} W = \Sigma_{U^T} \Delta_S' \Sigma_W^{-1} \Delta Y.$$

Define the matrix $\underline{T}$ by

$$\underline{T} = \Sigma_{U^T} \Delta_S' \Sigma_W^{-1} \Delta.$$

In a similar fashion,

$$\hat{V^T} = \Sigma_{V^T} \Delta_T' \Sigma_W^{-1} W = \Sigma_{V^T} \Delta_T' \Sigma_W^{-1} \Delta Y.$$

Now $\hat{T}_*$ is given by (4.10) of Bell and Hilmer (1988):

$$\begin{aligned} \hat{T}_* = &\left[1_{d_T} \;\; 0_{d_T \times d_S}\right] [H_1 \;\; H_2]^{-1} \\ &\times \left(Y_* - C_1 \left[1_{d_S} \;\; 0_{d_S \times n-d}\right] \hat{U^T} - C_2 \left[1_{d_T} \;\; 0_{d_T \times n-d}\right] \hat{V^T}\right). \end{aligned}$$

Here Y_* denotes the first d values of Y, i.e., $Y_* = [1_d \, 0]\, Y$. The matrices H_1 and H_2 are intimately involved in the initial value equations expounded in (3.2) of Bell (1984). In particular,

$$H_1 = \begin{bmatrix} 1_{d_T} \\ A_{d_T+1}^{T\prime} \\ \cdots \\ A_d^{T\prime} \end{bmatrix}$$

with the ith entry of the column vector A_j^T given by

$$\sum_{l=0}^{d_T-i} \left(-\delta_l^T\right) \xi_{j-i-l}^T$$

for $1 \leqslant i \leqslant d_T$. Also ξ_k^T denotes the kth coefficient of the polynomial $\xi^T(x)$, which is zero if k is negative. So H_1 has d rows and d_T columns. In a similar fashion, the d by d_S matrix H_2 is defined by

$$H_2 = \begin{bmatrix} 1_{d_S} \\ A_{d_S+1}^{S\prime} \\ \cdots \\ A_d^{S\prime} \end{bmatrix}$$

and the ith entry of A_j^S is

$$\sum_{l=0}^{d_S-i} \left(-\delta_l^S\right) \xi_{j-i-l}^S$$

for $1 \leqslant i \leqslant d_S$. Bell (1984) establishes that the matrix $[H_1 \;\; H_2]$ is invertible. The matrices C_1 and C_2 are d by d_S and d by d_T dimensional, respectively. Their entries are given by

$$C_{1ij} = \xi_{i-j-d_T}^T, \quad C_{2ij} = \xi_{i-j-d_S}^S.$$

Let J denote the matrix $[H_1 \;\; H_2]^{-1}$; then the expression for $\hat{T}_*$ can be simplified to

$$\begin{aligned} \hat{T}_* &= \left[1_{d_T} \;\; 0\right] J \left\{[1_d \;\; 0] - \left(C_1 \left[1_{d_S} \;\; 0\right] \Sigma_{U^T} \Delta_S' + C_2 \left[1_{d_T} \;\; 0\right] \Sigma_{W^S} \Delta_T'\right) \Sigma_W^{-1} \Delta\right\} Y \\ &= \overline{T} Y, \end{aligned}$$

where we define $\overline{T}$ to be the above d_T by n matrix. In an analogous derivation, letting $K = [H_2 \;\; H_1]^{-1}$,

$$\begin{aligned} \overline{S} &= \left[1_{d_S} \;\; 0\right] K \left\{[1_d \;\; 0] - \left(C_1 \left[1_{d_S} \;\; 0\right] \Sigma_{W^T} \Delta_S' + C_2 \left[1_{d_T} \;\; 0\right] \Sigma_{U^S} \Delta_T'\right) \Sigma_W^{-1} \Delta\right\} \\ &= \left[0 \;\; 1_{d_S}\right] J \left\{[1_d \;\; 0] - \left(C_1 \left[1_{d_S} \;\; 0\right] \Sigma_{W^T} \Delta_S' + C_2 \left[1_{d_T} \;\; 0\right] \Sigma_{U^S} \Delta_T'\right) \Sigma_W^{-1} \Delta\right\}. \end{aligned}$$

Also let

$$\underline{S} = \Sigma_{U^S} \Delta_T' \Sigma_W^{-1} \Delta.$$

The next result follows from the above formulas:

Proposition 2. *For the full model under Assumption A, the signal extraction matrices are given by*

$$F_{STI}^T = \overline{\Delta}_T^{-1} \begin{bmatrix} Q_T \overline{T} \\ \underline{T} \end{bmatrix}, \quad F_{STI}^S = \overline{\Delta}_S^{-1} \begin{bmatrix} Q_S \overline{S} \\ \underline{S} \end{bmatrix}.$$

Remark 2. In principle, these above formulas are sufficient to generate optimal (in a minimal mean squared error sense) linear signal extraction estimates under Assumption A. However, the formulas for $\overline{S}$ and $\overline{T}$, which produce initial value estimates for the signals, are quite complicated; in contrast, signal extraction for the reduced models is simpler. In Section 4, an iterative method is developed to produce the full model signal extraction matrices without explicit recourse to the initial value matrices $\overline{S}$ and $\overline{T}$.

3. Formulas relating the full model filters to the reduced model filters

For the remainder of this paper, we only assume Assumption A for the full model, which determines the expectation operator $\mathbb{E}$. The reduced model filters are defined by Eq. (4); they are optimal among linear estimates whose error does not depend on initial values. Hence (1) is not valid for the $\mathbb{E}$ operator defined for the full model, even though the reduced model filters are defined as if (1) were true.

Below, we will need to examine the eigenvalues of $F_{SI}^S F_{TI}^T$. For notation, we let $\lambda_1(A), \ldots, \lambda_n(A)$ denote the eigenvalues of a matrix A in descending order. The following proposition summarizes some important properties of this matrix.

Proposition 3. *Define F_{SI}^S and F_{TI}^T as in Proposition 1. We have*

$$0 < \lambda_n\left(F_{SI}^S F_{TI}^T\right) \leqslant \lambda_1\left(F_{SI}^S F_{TI}^T\right) < 1.$$

Also the inverse of $1_n - F_{SI}^S F_{TI}^T$ exists. The series $\sum_{k=0}^{\infty}\left(F_{SI}^S F_{TI}^T\right)^k$ is convergent with sum equal to $\left(1_n - F_{SI}^S F_{TI}^T\right)^{-1}$. The same results hold for $F_{TI}^T F_{SI}^S$ in place of $F_{SI}^S F_{TI}^T$.

Proof. We first show the invertibility of $1_n - F_{SI}^S F_{TI}^T$. Observe that

$$\begin{aligned}
&1_n - F_{SI}^S F_{TI}^T \\
&\quad = F_{SI}^S\left(\left(F_{SI}^S\right)^{-1}\left(F_{TI}^T\right)^{-1} - 1_n\right)F_{TI}^T \\
&\quad = F_{SI}^S\left(\Sigma_I \Delta_S^{*\prime}\Sigma_{U^S}^{-1}\Delta_S^* + \Sigma_I \Delta_T^{*\prime}\Sigma_{U^T}^{-1}\Delta_T^* + \Sigma_I \Delta_S^{*\prime}\Sigma_{U^S}^{-1}\Delta_S^*\Sigma_I \Delta_T^{*\prime}\Sigma_{U^T}^{-1}\Delta_T^*\right)F_{TI}^T.
\end{aligned}$$

If the central matrix in parentheses is invertible, then so is $1_n - F_{SI}^S F_{TI}^T$. We begin by computing the minimal eigenvalue of $\left(F_{SI}^S\right)^{-1}\left(F_{TI}^T\right)^{-1}$. Note that Σ_I is positive definite, so its square root is well-defined. Using $\lambda(A) = \lambda\left(\Sigma_I^{-1/2} A \Sigma_I^{1/2}\right)$, we obtain

$$\begin{aligned}
&\lambda_n\left(1_n + \Sigma_I \Delta_S^{*\prime}\Sigma_{U^S}^{-1}\Delta_S^* + \Sigma_I \Delta_T^{*\prime}\Sigma_{U^T}^{-1}\Delta_T^* + \Sigma_I \Delta_S^{*\prime}\Sigma_{U^S}^{-1}\Delta_S^*\Sigma_I \Delta_T^{*\prime}\Sigma_{U^T}^{-1}\Delta_T^*\right) \\
&\quad = \lambda_n\left(\left(1_n + \Sigma_I^{1/2\prime}\Delta_S^{*\prime}\Sigma_{U^S}^{-1}\Delta_S^*\Sigma_I^{1/2}\right)\left(1_n + \Sigma_I^{1/2\prime}\Delta_T^{*\prime}\Sigma_{U^T}^{-1}\Delta_T^*\Sigma_I^{1/2}\right)\right) \\
&\quad = \lambda_n(GH).
\end{aligned}$$

Both of these matrices G and H are symmetric, and it is easy to check that their minimal eigenvalues are $\geqslant 1$, so that they are positive definite too. Hence they have a Cholesky factorization, and

$$\lambda_n(GH) = \lambda_n\left(G^{1/2'}HG^{1/2}\right) = \inf_{x\neq 0} \frac{x'G^{1/2'}HG^{1/2}x}{x'x} = \inf_{y\neq 0} \frac{y'Hy}{y'G^{-1}y}.$$

We can compute the inverse of G:

$$\begin{aligned} G^{-1} &= \left(1_n + \Sigma_I^{1/2'}\Delta_S^{*'}\Sigma_{U^S}^{-1}\Delta_S^*\Sigma_I^{1/2}\right)^{-1} \\ &= \Sigma_I^{-1/2}\left(1_n + \Sigma_I\Delta_S^{*'}\Sigma_{U^S}^{-1}\Delta_S^*\right)^{-1}\Sigma_I^{1/2} \\ &= \Sigma_I^{-1/2}\left(1_n - \Sigma_I\Delta_S^{*'}\Sigma_{W^S}^{-1}\Delta_S^*\right)\Sigma_I^{1/2} \\ &= 1_n - \Sigma_I^{1/2'}\Delta_S^{*'}\Sigma_{W^S}^{-1}\Delta_S^*\Sigma_I^{1/2}. \end{aligned}$$

Now the minimal eigenvalue of G^{-1} is the reciprocal of the maximum eigenvalue of G, which is $\geqslant 1$. So $\lambda_n\left(G^{-1}\right) > 0$, and this in turn implies that

$$x'\Sigma_I^{-1}x - x'\Delta_S^{*'}\Sigma_{W^S}^{-1}\Delta_S^*x > 0$$

for all $x \neq 0$. Hence

$$\lambda_n(GH) = \inf_{z\neq 0} \frac{z'\Sigma_I^{-1}z + z'\Delta_T^{*'}\Sigma_{U^T}^{-1}\Delta_T^*z}{z'\Sigma_I^{-1}z - z'\Delta_S^{*'}\Sigma_{W^S}^{-1}\Delta_S^*z}$$

is nonnegative and bounded. Now this quantity, using the nonnegative definiteness of $\Delta_T^{*'}\Sigma_{U^T}^{-1}\Delta_T^*$ and $\Delta_S^{*'}\Sigma_{W^S}^{-1}\Delta_S^*$, is at least one, and is equal to one if and only if z is in the null space of Δ_T^* and Δ_S^*. This only happens if $z=0$, as the following lemma demonstrates, and hence $\lambda_n(GH) > 1$.

Lemma 2. *If δ^T and δ^S share no common roots, then the intersection of the null spaces of Δ_T^* and Δ_S^* is zero.*

The proof is in the appendix. Now since $\lambda_n(GH) > 1$, we have

$$\lambda_n\left(\left(F_{SI}^S\right)^{-1}\left(F_{TI}^T\right)^{-1} - 1\right) = \lambda_n(GH) - 1 > 0$$

which implies that this matrix is invertible. Hence $\left(1 - F_{SI}^S F_{TI}^T\right)^{-1}$ exists.

For the first assertion of the proposition, the inequality involving the eigenvalues of $F_{SI}^S F_{TI}^T$, we consider

$$\lambda_j\left(F_{SI}^S F_{TI}^T\right) = \lambda_j\left(G^{-1}H^{-1}\right)$$

for any j. Letting $j=1$, and using the positive definiteness of G^{-1}, we obtain

$$\lambda_1\left(G^{-1}H^{-1}\right)=\lambda_1\left(G^{-1/2}H^{-1}G^{-1/2}\right)=\frac{1}{\lambda_n\left(G^{1/2}HG^{1/2}\right)}=\frac{1}{\lambda_n(GH)}<1.$$

Similarly $\lambda_n\left(G^{-1}H^{-1}\right)\geqslant\lambda_n\left(G^{-1}\right)\lambda_n\left(H^{-1}\right)>0$ by Corollary 3.14 of Axelsson (1996).

Finally, we compute the Schur decomposition of $F^S_{SI}F^T_{TI}$ as follows:

$$Q'F^S_{SI}F^T_{TI}Q=\Lambda+N$$

for Q orthogonal, Λ a diagonal matrix with the eigenvalues as entries, and N strictly upper triangular. Then by Lemma 7.3.2 of Golub and Van Loan (1996), we can choose any $\theta\geqslant 0$ and obtain the bound

$$\left\|\left(F^S_{SI}F^T_{TI}\right)^k\right\|_2\leqslant(1+\theta)^{n-1}\left(\lambda_1\left(F^S_{SI}F^T_{TI}\right)+\|N\|_F/(1+\theta)\right)^k, \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Set

$$\theta=\frac{2\|N\|_F}{1-\lambda_1\left(F^S_{SI}F^T_{TI}\right)}-1$$

and then

$$\lambda_1\left(F^S_{SI}F^T_{TI}\right)+\|N\|_F/(1+\theta)=\tfrac{1}{2}\left(1+\lambda_1\left(F^S_{SI}F^T_{TI}\right)\right)<1.$$

Denote this quantity by η, and let $C=(1+\theta)^{n-1}$. Then the partial sums of the powers of $F^S_{SI}F^T_{TI}$ are a Cauchy sequence:

$$\left\|\sum_{k=M}^{M+L}\left(F^S_{SI}F^T_{TI}\right)^k\right\|_2\leqslant\sum_{k=M}^{M+L}\left\|\left(F^S_{SI}F^T_{TI}\right)^k\right\|_2=C\eta^M\frac{1-\eta^{L+1}}{1-\eta}$$

for any positive integers M and L. This clearly tends to zero as M and L tend to infinity, and hence the series $\sum_{k\geqslant 0}\left(F^S_{SI}F^T_{TI}\right)^k$ converges. Also letting $L=0$, we see that $\left(F^S_{SI}F^T_{TI}\right)^k\to 0$. So taking the limit as $n\to\infty$ in

$$\sum_{k=0}^{n}\left(F^S_{SI}F^T_{TI}\right)^k\left(1-F^S_{SI}F^T_{TI}\right)=1-\left(F^S_{SI}F^T_{TI}\right)^{n+1}$$

shows that $\sum_{k\geqslant 0}\left(F^S_{SI}F^T_{TI}\right)^k=\left(1-F^S_{SI}F^T_{TI}\right)^{-1}$. $\square$

If all three components were stationary, we would easily see that

$$F^T_{STI}=F^T_{TI}\left(1_n-F^S_{STI}\right)$$

since $F_{TI}^{T} = \Sigma_T \Sigma_{TI}^{-1}$, $F_{STI}^{T} = \Sigma_T \Sigma_Y^{-1}$, and $F_{STI}^{S} = \Sigma_S \Sigma_Y^{-1}$, and by orthogonality of components, it follows that $\Sigma_Y = \Sigma_T + \Sigma_S + \Sigma_I$. The proof of the above relation is then

$$\begin{aligned} F_{TI}^{T}\left(1_n - F_{STI}^{S}\right) &= \Sigma_T (\Sigma_T + \Sigma_I)^{-1}\left(1_n - \Sigma_S \Sigma_Y^{-1}\right) \\ &= \Sigma_T (\Sigma_Y - \Sigma_S)^{-1} (\Sigma_Y - \Sigma_S) \Sigma_Y^{-1} \\ &= \Sigma_T \Sigma_Y^{-1} = F_{STI}^{T}. \end{aligned}$$

This result also holds when the trend and seasonal are nonstationary, as the following theorem demonstrates.

Theorem 1. *Given the definitions of the* F_{TI}^{T} *and* F_{SI}^{S} *matrices in* (4), *suppose that the assumptions made in the beginning of Section* 2 *are true. Also, suppose that Assumption A holds for the stochastic process Y. Then the following formulas hold*:

$$\begin{aligned} F_{STI}^{T} &= F_{TI}^{T}\left(1_n - F_{STI}^{S}\right), \\ F_{STI}^{S} &= F_{SI}^{S}\left(1_n - F_{STI}^{T}\right). \end{aligned} \tag{6}$$

These equations can be solved simultaneously to yield

$$\begin{aligned} \left(1_n - F_{TI}^{T} F_{SI}^{S}\right) F_{STI}^{T} &= F_{TI}^{T}\left(1_n - F_{SI}^{S}\right), \\ \left(1_n - F_{SI}^{S} F_{TI}^{T}\right) F_{STI}^{S} &= F_{SI}^{S}\left(1_n - F_{TI}^{T}\right). \end{aligned} \tag{7}$$

In addition, the matrices $\left(1_n - F_{SI}^{S} F_{TI}^{T}\right)$ *and* $\left(1_n - F_{TI}^{T} F_{SI}^{S}\right)$ *are invertible, which results in the explicit formulas*

$$\begin{aligned} F_{STI}^{T} &= \left(1_n - F_{TI}^{T} F_{SI}^{S}\right)^{-1} F_{TI}^{T}\left(1_n - F_{SI}^{S}\right), \\ F_{STI}^{S} &= \left(1_n - F_{SI}^{S} F_{TI}^{T}\right)^{-1} F_{SI}^{S}\left(1_n - F_{TI}^{T}\right). \end{aligned} \tag{8}$$

Remark 3. The first pair of formulas (6) give an intuitive interpretation of the relationship between F_{STI}^{T} and F_{STI}^{S}. For example, trend extraction is actually the same as seasonal adjustment followed by trend extraction for a "perfectly deseasonalized" series. Likewise, seasonal extraction is detrending followed by seasonal estimation for a detrended series. The latter formulas (8) express these full model filters entirely in terms of reduced model filters.

Proof. Define $C_t = T_t + S_t$; then if we wish to extract the nonstationary signal C from the full model Y, we have the simple formula (since the noise I is stationary):

$$F_{STI}^{C} = 1_n - \Sigma_I \Delta' \Sigma_W^{-1} \Delta$$

which is derived in a similar fashion to the matrices in Proposition 1. Now we wish to show that

$$F_{STI}^{C} = F_{STI}^{T} + F_{STI}^{S}. \tag{9}$$

These three matrices are defined under Assumption A, and thus the first and last equalities of

$$F^C_{STI}Y = E(C|Y) = E(T+S|Y) = E(T|Y) + E(S|Y) = F^T_{STI}Y + F^S_{STI}Y$$

are valid for all Y; hence (9) must hold. From (9) we can write

$$\begin{aligned} F^T_{STI} &= \left(1_n - \Sigma_I \Delta' \Sigma_W^{-1} \Delta\right) - F^S_{STI} \\ &= -\Sigma_I \Delta' \Sigma_W^{-1} \Delta + \left(1_n - F^S_{STI}\right). \end{aligned}$$

The rest of the proof of the formula is a calculation:

$$\begin{aligned} & F^T_{TI}\left(1 - F^S_{STI}\right) \\ &= F^T_{TI}\left(F^T_{STI} + \Sigma_I \Delta' \Sigma_W^{-1} \Delta\right) \\ &= F^T_{STI} - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* F^T_{STI} - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I \Delta' \Sigma_W^{-1} \Delta + \Sigma_I \Delta' \Sigma_W^{-1} \Delta \\ &= F^T_{STI} - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Sigma_{U^T} \Delta_S' \Sigma_W^{-1} \Delta - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I \Delta' \Sigma_W^{-1} \Delta + \Sigma_I \Delta' \Sigma_W^{-1} \Delta \\ &= F^T_{STI} - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \left(\Sigma_{U^T} + \Delta_T^* \Sigma_I \Delta_T^{*\prime} - \Sigma_{W^T}\right) \Delta_S' \Sigma_W^{-1} \Delta = F^T_{STI} \end{aligned}$$

which uses the fact that

$$\Delta_T^* F^T_{STI} = \left[0 1_{n-d_T}\right] \overline{\Delta}_T \overline{\Delta}_T^{-1} \begin{bmatrix} Q_T \overline{T} \\ \underline{T} \end{bmatrix} = \underline{T} = \Sigma_{U^T} \Delta_S' \Sigma_W^{-1} \Delta.$$

The second equation in (6) has a similar proof.

Now if we solve the pair of equations in (6) we immediately obtain (7). The invertibility of $1 - F^S_{SI} F^T_{TI}$ is demonstrated in Proposition 3. With analogous calculations for the second line of (8), the proof is complete. □

4. The main algorithm and its analysis

The idea of the algorithm of this section is to use (6) to define an iteration scheme. The resulting algorithm produces estimates of signal and trend $\hat{S}$ and $\hat{T}$ that satisfy

$$\begin{aligned} \hat{T} &= F^T_{TI}(Y - \hat{S}), \\ \hat{S} &= F^S_{SI}(Y - \hat{T}) \end{aligned}$$

which is essentially (6) applied to Y. This is done essentially by constructing the solutions to the linear system defined by applying (8) to the vector Y. These signal extraction estimates are the unique conditional expectation estimates under Assumption A, and are still locally optimal if Assumption A is not true, as discussed in the introduction. A nice feature of the algorithm is its idempotency, i.e., if one inputs the limiting values into the algorithm as initial values, the same limiting values are returned unaltered as output. The algorithm converges quickly; below, a geometric convergence rate is derived. Any initialization of the algorithm can be used.

Theorem 2. *Consider the following algorithm*:

$$
\begin{aligned}
&\hat{S}^{(0)} \quad \textit{is any given vector} \\
&\textit{for } i = 1 \textit{ to convergence} \\
&\quad \hat{T}^{(i)} = F^T_{TI}\left(Y - \hat{S}^{(i-1)}\right) \\
&\quad \hat{S}^{(i)} = F^S_{SI}\left(Y - \hat{T}^{(i)}\right) \\
&\textit{end for}
\end{aligned}
$$

The algorithm converges geometrically fast to $\hat{S}^{(\infty)} = F^S_{STI}Y$ *and* $\hat{T}^{(\infty)} = F^T_{STI}Y$.

Proof. We will analyze the iterations of $\hat{S}^{(i)}$. Simple algebra produces

$$\hat{S}^{(i+1)} = F^S_{SI}\left(1_n - F^T_{TI}\right)Y + F^S_{SI}F^T_{TI}\hat{S}^{(i)}$$

from which it follows that

$$\hat{S}^{(i+1)} - \hat{S}^{(i)} = F^S_{SI}\left(1_n - F^T_{TI}\right)Y - \left(1_n - F^S_{SI}F^T_{TI}\right)\hat{S}^{(i)}.$$

At this stage, note that if $\hat{S}^{(i)} = F^S_Y Y$, then by (7) we have

$$\hat{S}^{(i+1)} - \hat{S}^{(i)} = F^S_{SI}\left(1_n - F^T_{TI}\right)Y - \left(1_n - F^S_{SI}F^T_{TI}\right)F^S_{STI}Y = 0$$

so that this is a fixed point of the mapping. One interesting aspect of this is "double idempotency," i.e., if $\hat{S}^{(i)} = F^S_{STI}Y$, then automatically $\hat{T}^{(i+1)} = F^T_{STI}Y$ as well. Now through a simple induction on i we obtain

$$\hat{S}^{(i+1)} = \sum_{j=0}^{i}\left(F^S_{SI}F^T_{TI}\right)^j F^S_{SI}\left(1_n - F^T_{TI}\right)Y + \left(F^S_{SI}F^T_{TI}\right)^{i+1}\hat{S}^{(0)}.$$

Similarly, we can compute the iterate of the trend to be

$$\hat{T}^{(i+1)} = \sum_{j=0}^{i}\left(F^T_{TI}F^S_{SI}\right)^j F^T_{TI}\left(1_n - F^S_{SI}\right)Y + \left(F^T_{TI}F^S_{SI}\right)^i F^T_{TI}\left(F^S_{SI}Y - \hat{S}^{(0)}\right).$$

Now, using the fact that $\sum_{j\geqslant 0}\left(F^T_{TI}F^S_{SI}\right)^j$ and $\sum_{j\geqslant 0}\left(F^S_{SI}F^T_{TI}\right)^j$ are convergent, as shown in Proposition 3, and the fact that $\left(F^T_{TI}F^S_{SI}\right)^i$ and $\left(F^S_{SI}F^T_{TI}\right)^i$ tend to zero as $i \to \infty$, we see that the iterates converge to

$$\begin{aligned}\hat{S}^{(\infty)} &= \sum_{j=0}^{\infty} \left(F_{SI}^{S} F_{TI}^{T}\right)^{j} F_{SI}^{S} \left(1_n - F_{TI}^{T}\right) Y \\ &= \left(1_n - F_{SI}^{S} F_{TI}^{T}\right)^{-1} F_{SI}^{S} \left(1_n - F_{TI}^{T}\right) Y = F_{STI}^{S} Y, \\ \hat{T}^{(\infty)} &= \sum_{j=0}^{\infty} \left(F_{TI}^{T} F_{SI}^{S}\right)^{j} F_{TI}^{T} \left(1_n - F_{SI}^{S}\right) Y \\ &= \left(1_n - F_{TI}^{T} F_{SI}^{S}\right)^{-1} F_{TI}^{T} \left(1_n - F_{SI}^{S}\right) Y = F_{STI}^{T} Y\end{aligned}$$

independent of the initialization $\hat{S}^{(0)}$. As for the rate of convergence, the difference between successive iterates will decay at geometric rate, as shown in relation (5). □

Remark 4. It is interesting that the algorithm gradually computes the inverse of $1_n - F_{TI}^{T} F_{SI}^{S}$, along with a decaying error matrix that multiplies the initialization $\hat{S}^{(0)}$. This algorithm has been implemented in the Ox language and tested on airline model decompositions —we performed the canonical decomposition into trend, seasonal and irregular in the sense of Hilmer and Tiao (1982) on a Box–Jenkins airline model. In most cases the estimates had essentially converged by the third iteration. This convergence can be slowed down by erratic choices of $\hat{S}^{(0)}$, such as a white noise sequence with high variance, but the proof of Theorem 2 shows that the initialization has no effect in the long term. For most applications, one would take $\hat{S}^{(0)}$ to be the zero vector—a "noninformative" choice.

5. Computer implementation

This section contains a short discussion of the computer implementation of the main algorithm. First we present two examples of Theorem 2 in action. We simulated a monthly series of length 49 from the airline model (3) with parameters $\theta = 0.6$, $\Theta = 0.4$, and innovation variance 1. The series was initialized with 13 values from a real series (though these were not plotted) exhibiting locally linear trend and stochastic seasonality. The algorithm was initialized with $\hat{S}^{(0)} = 0$ the zero vector. For the simulated series, θ, Θ, and the innovation variance were then estimated (since these determine the filters, and the estimated values can differ significantly from truth, it is important to estimate). Once the full model is known, we used the canonical decomposition approach of Hilmer and Tiao (1982) to obtain the component models. Then we explicitly computed the reduced model filter formulas from these model parameters. For the full model filters, which we obtained in order to check our results, we used the Kalman filter and smoother of SsfPack (Koopman et al., 1999). Then we implemented the algorithm directly with the computed filter matrices.

The algorithm converged after seven iterations, where convergence was measured by whether the vector two-norm of successive seasonal and trend iterates was less than 0.1. Notice that in earlier iterations, some seasonality is present in the trend estimates, and

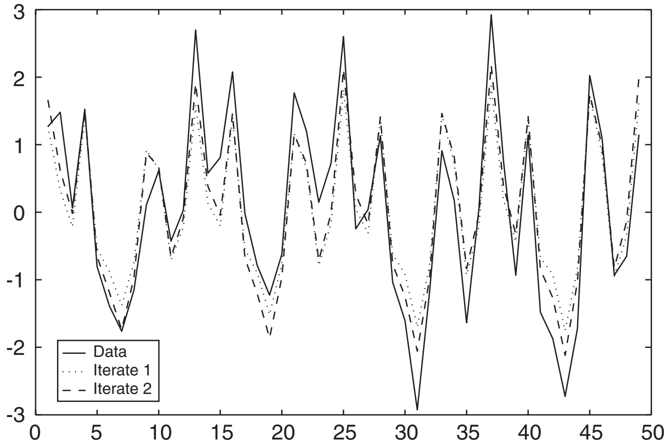

Fig. 1. Seasonal iterates.

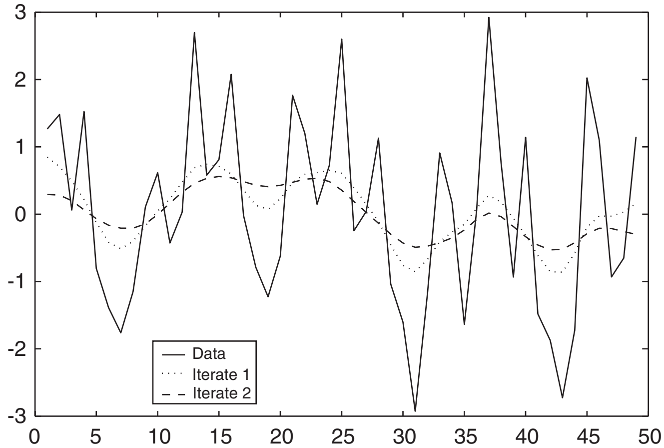

Fig. 2. Trend iterates.

some trend is in the seasonal estimates, but this confusion of signals is gradually weeded out—compare Figs. 1 and 2 with Fig. 3. In fact, an examination of the filter weights at the center of the sample (Figs. 4 and 5) shows that the reduced model trend filters are somewhat shorter (i.e., more of their weights are close to zero) than the full model filters. Not only is the full model trend filter a bit longer than the reduced, but one can easily see the seasonal suppression that it performs, which is the visual analogue of (6). This simulation example was chosen to demonstrate a situation where only weak seasonality is present—this allows one to visualize the convergence of the seasonal iterates. Typically, simulation with more distinct seasonality produced seasonal iterates that had essentially converged by the first iteration.

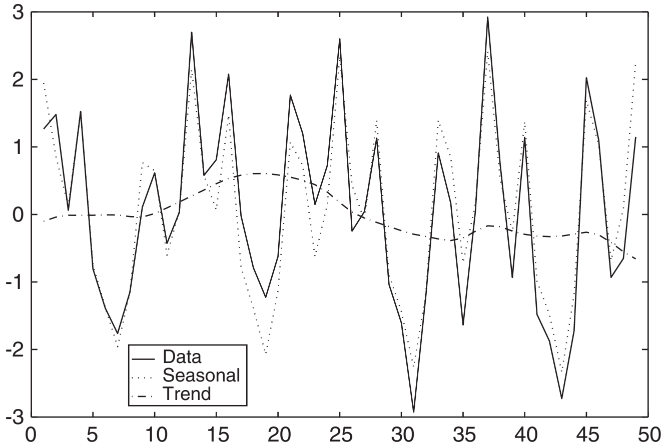

Fig. 3. Optimal estimates.

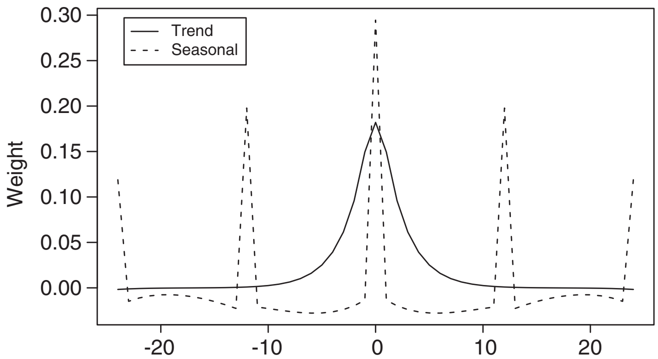

Fig. 4. Reduced model weights.

Next, we analyzed the US Retail Sales of Shoe Stores data from the Monthly Retail Trade Survey, from 1984 to 1998. After adjusting for outliers using X-12 ARIMA, the logged data were fitted to an airline model as in the simulation study. This produced values of $\theta = 0.572$ and $\Theta = 0.336$, and innovation standard deviation 0.031. The algorithm was initialized with $\hat{S}^{(0)} = 0$ the zero vector, and under the same convergence criterion, it converged in 10 iterations. Figs. 6 and 7 show the seasonal and trend iterates for the first 5 years of data, together with the final estimates. In comparing the squared gains (Figs. 8 and 9) for the reduced and full model filters, one can again see the seasonal suppression of the full model trend filter, whereas the reduced model gains are comparatively simpler.

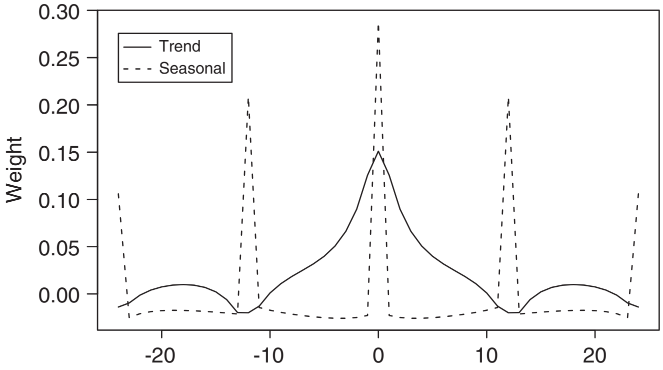

Fig. 5. Full model weights.

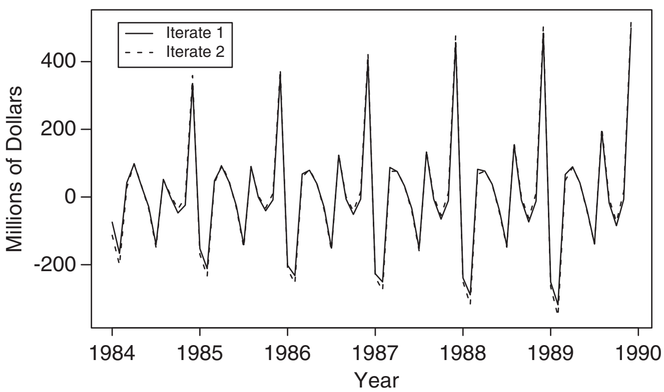

Fig. 6. Seasonal iterates for shoe sales.

5.1. Notes on implementation

These plots were produced through Ox code—see Doornik (1998). One may use a State Space Representation and SsfPack (Koopman et al., 1999) to produce the various filters, but this is not necessary for the reduced model filters, since their formulas are so simple. The basic portions of the program were: simulation, estimation, decomposition, filter construction, application of the algorithm, and visualization of the results. For simulation, we note that the initial values can have a significant impact on the data generated; this in turn can affect the estimated parameter values and thereby change the filters. It is necessary to obtain models for the S, T, and I components, and there are two popular choices. The Structural Models approach estimates the model parameters for the components directly from the data. The canonical decomposition approach is different in that it first estimates a model for the full data $Y = S + T + I$, and then analytically computes component models such that the irregular innovation variance is maximized—see Hilmer and Tiao (1982). In

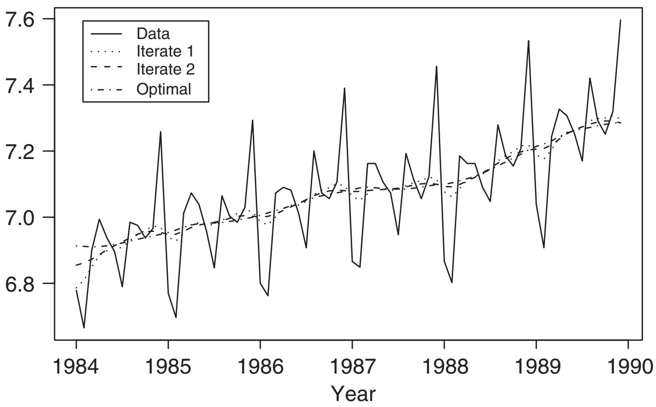

Fig. 7. Trend iterates for shoe sales.

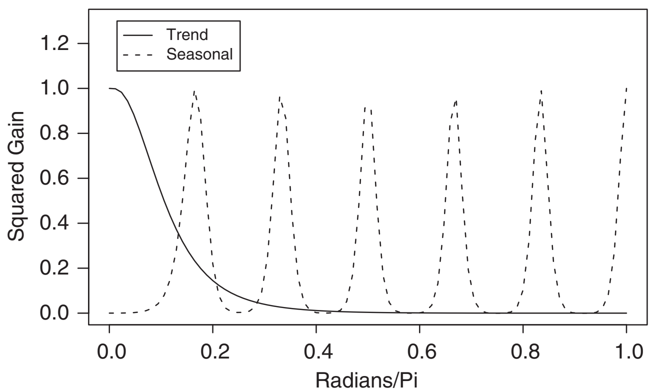

Fig. 8. Reduced model squared gain.

this paper we have followed the latter approach, though the second author has done some implementations with Structural Models as well.

Once models for the components are known, we can apply the results of this paper. To compute the filters F^T_{TI} and F^S_{SI} it is necessary to know, by Proposition 1, Σ_I, Σ_{U^T}, and Σ_{U^S}. These are simply the Toeplitz autocovariance matrices of the irregular, differenced trend, and differenced seasonal—hence they are easily obtained from the component models. We chose to use the first formula in Proposition 1 so that only one matrix inversion would be necessary. Alternatively, these matrices can be produced automatically by software that does Kalman smoothing, such as SsfPack. Once the filters have been computed, we simply apply the algorithm. Again, in our implementation in Ox we produced the full filters F^T_{STI} and F^S_{STI} from SsfPack to check the results of the algorithm—the two methods produced identical matrices (up to the error inherent in iterating our algorithm only a finite number of steps).

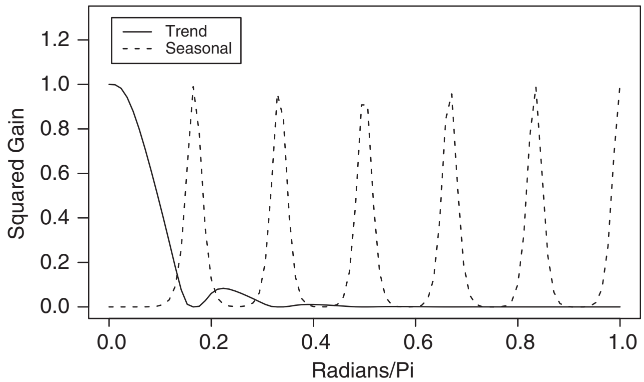

Fig. 9. Full model squared gain.

6. Conclusion

This paper has linked the trend and seasonal extraction matrices when these components are nonstationary. Of course, Theorem 1 will apply to any three component model where at least one of the components is stationary, e.g., trend plus cycle plus irregular econometric models. The algorithm of Section 4 presents a method for building up the correct signal extraction filters from less complicated, more intuitive (and more commonly used) reduced model filters.

6.1. Comparison with other methods

There are several ways of going about model-based signal extraction these days, and we will briefly discuss a few of them here. The SEATS program of Gómez and Maravall (1994) takes the canonical decomposition approach to generate component models, and then applies the bi-infinite signal extraction filters of Bell (1984) to forecast-extended data, which actually results in the finite sample signal extraction estimates. Alternatively, the Kalman smoother can be used to generate the extractions from the component models, and has the advantage that the finite sample Mean Squared Errors are also produced—see Koopman et al. (1999). These authors' software, SsfPack, can currently deal with structural models, i.e., estimate the component models using a structural approach, but a future version of SsfPack will do canonical decompositions as well. The method of this paper provides a third approach, although all three generate the same estimates under Assumption A on the data. Our technique is simple to explain and implement, and this simplicity makes it fast. The key advantage of the Kalman smoother is its versatility—not only the wealth of problems it can handle, but the diversity of applications it produces. SEATS is unable to produce finite sample Mean Squared Errors, in contrast to the Kalman smoother. Our approach goes a step further—it is simple to compute the entire covariance matrix of signal extraction errors (not merely the diagonal of this matrix) from our formulas. These error formulas—given by

$$\left(1_n - F^T_{TI} F^S_{SI}\right)^{-1} F^T_{TI} \Sigma_I,$$
$$\left(1_n - F^S_{SI} F^T_{TI}\right)^{-1} F^S_{SI} \Sigma_I$$

for the trend and seasonal extraction error covariances, respectively—are useful for determining the variance of statistics computed from estimated signals.

6.2. Analogy with X-11

As mentioned earlier in the paper, our procedure is similar in spirit to the main algorithm of X-11. Leaving aside the extreme value adjustment, the additive X-11 program starts by constructing a crude trend, using nonparametric filters that implicitly assume a "trend plus noise" form of the data. Thus, this is analogous to our estimation of $\hat{T}^{(1)}$, the key difference being that our procedure is model-based whereas X-11 is nonparametric. Next, X-11 produces a seasonal estimate from the detrended data, i.e., the data with crude trend subtracted. This is analogous to our estimation of $\hat{S}^{(1)}$. X-11 then iterates this whole process one more time, producing analogies of $\hat{T}^{(2)}$ and $\hat{S}^{(2)}$. This process is done within each of the B, C, and D iterations of X-11 (see Ladiray and Quenneville (2001)). Hence, we may be justified in viewing our procedure as a model-based version of X-11 without extreme value adjustment and with additional iterations (and initializations) possible. If one now inserts an extreme value adjustment procedure between each iteration—for example, X-11's extreme value adjustment process—then the analogy is even closer. This combination of model-based signal extraction with the rugged stability of X-11 is appealing and powerful.

6.3. Connections to common practice

Finally, we observe that many practitioners may be essentially using the first iteration of this paper's algorithm, with the filters F^T_{TI} and F^S_{SI}. In the model-based analogue of the first iteration of the additive X-11 algorithm, we first estimate trend with a trend extraction matrix for the two-component trend plus irregular model—we are essentially applying F^T_{TI} to the data; if we then detrend, we have

$$Y - F^T_{TI} Y.$$

Then this would be followed up with a seasonal extraction filter for a reduced two component model, which is F^S_{SI}. Our resultant seasonal estimate is

$$F^S_{SI}\left(1_n - F^T_{TI}\right) Y$$

which is the first iteration of our algorithm with an initial value of $\hat{S}^{(0)} = 0$. Similarly, in the arena of cycle estimation, one typically first detrends with a low-pass filter and follows up with a high-pass or band-pass filter to extract the cycle—see Harvey and Trimbur (2003). In this case, one could conceptually replace the seasonal S above by a cycle component C. Again, this would be step one of our algorithm with an initialization of zero (here replacing the seasonal with a cycle). Simulation work seems to indicate that at least two or three iterations, rather than one, should be performed in order to get reasonably close to the optimal estimates. This paper suggests that these practitioners will obtain better results by iterating their entire signal extraction procedure a few times.

Acknowledgements

The first author would like to thank David Findley for communicating the main questions of this paper; he is also grateful to John Aston for pointing out references for the matrix algebra, and for writing portions of the Ox code used to implement the algorithm of Section 4. He thanks Thomas Trimbur for discussions on trend-cycle estimation in the econometrics literature, as well as Bill Bell for useful comments. Thanks also go to the referees and Stephen Pollock for many useful suggestions.

Appendix A. Technical proof

Proof of Lemma 2. The equation

$$\Delta_T^* x = 0$$

involves a choice of d_T initial values for x, with the subsequent components determined by the difference equation

$$-\delta_{d_T}^T x_k - \delta_{d_T-1}^T x_{k+1} - \cdots - \delta_1^T x_{k+d_T-1} - \delta_0^T x_{k+d_T} = 0.$$

If we denote the h distinct roots of δ^T by z_i, which each have multiplicity m_i, then according to page 586 of Henrici (1988) the general solution is a linear combination of

$$x_t^{(i,j)} = t^j z_i^{-t}$$

for $i = 1, \ldots, h$ and $j = 0, 1, \ldots, m_i - 1$; $t = 1, \ldots, n$. Note that $d_T = \sum_{i=1}^h m_i$, so the indices (i, j) specify a basis for all d_T solutions by Henrici (1988). Now the solutions to

$$\Delta_S^* y = 0$$

have the form

$$y_t^{(k,l)} = t^l w_k^{-t}$$

for w_k the roots of δ^S. Now fix (i, j) and (k, l), and let α and β be constants such that

$$0 = \alpha x_t^{(i,j)} + \beta y_t^{(k,l)} = \alpha t^j z_i^{-t} + \beta t^l w_k^{-t}$$

for all $t = 1, 2, \ldots, n$. Hence

$$0 = \alpha t^{j-l}(w_k/z_i)^t + \beta$$

which implies that either $\alpha = 0 = \beta$ or

$$\beta/\alpha = -t^{j-l}(w_k/z_i)^t.$$

But since $w_k \neq z_i$ for all i and k, this ratio must depend on t, implying that either α or β depends on t, an absurdity. Hence $\alpha=0=\beta$, so that the basis vectors are linearly independent. Hence only the zero vector can lie in the null spaces of both Δ_T^* and Δ_S^*. $\square$

References

Anderson, B., Moore, J., 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.

Ansley, C., Kohn, R., 1985. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. Ann. Statist. 13, 1286–1316.

Axelsson, O., 1996. Iterative Solution Methods. Cambridge University Press, Cambridge.

Bell, W., 1984. Signal extraction for nonstationary time series. Ann. Statist. 12, 646–664.

Bell, W., 2004. On RegComponent time series models and their applications. In: Harvey, A.C., Koopman, S.J., Shephard, N. (Eds.), State Space and Unobserved Component Models: Theory and Applications. Cambridge University Press, Cambridge, UK.

Bell, W., Hilmer, S., 1988. A matrix approach to likelihood evaluation and signal extraction for ARIMA component time series models. SRD Research Report No. *RR*-88/22, Bureau of the Census, available at http://www.census.gov/srd/www/byname.html.

Bell, W., Hilmer, S., 1991. Initializing the Kalman filter for nonstationary time series models. J. Time Ser. Anal. 12, 283–300.

Cleveland, W., Tiao, G., 1976. Decomposition of seasonal time series: a model for the census X-11 program. J. Amer. Statist. Assoc. 71, 581–587.

De Jong, P., 1991. The diffuse Kalman filter. Ann. Statist. 19, 1073–1083.

Doornik, J., 1998. Object-oriented Matrix Programming Using Ox 2.0. Timberlake Consultants Press, London.

Durbin, J., Koopman, S., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

Golub, G., Van Loan, C., 1996. Matrix Computations. The Johns Hopkins University Press, Baltimore and London.

Gómez, V., Maravall, A., 1994. Program SEATS signal extraction in ARIMA time series: instructions for the user. Working Paper ECO 94/28, European University Institute, Florence.

Hannan, E.J., 1967. Measurement of a wandering signal amid noise. J. Appl. Probab. 4, 90–102.

Harvey, A., Trimbur, T., 2003. General model-based filters for extracting cycles and trends in economic time series. Rev. Econom. Statist. 85 (2), 244–255.

Henrici, P., 1988. Applied and Computational Complex Analysis, vol. 1. Wiley, New York.

Hilmer, S., Tiao, G., 1982. An ARIMA-model-based approach to seasonal adjustment. J. Amer. Statist. Assoc. 77, 377, 63–70.

Hodrick, R.J., Prescott, E.C., 1997. Postwar US business cycles: an empirical investigation. J. Money Credit Banking 24, 1–16.

Kohn, R., Ansley, C.F., 1986. Estimation, prediction, and interpolation for ARIMA models with missing data. J. Amer. Statist. Assoc. 81, 751–761.

Kohn, R., Ansley, C.F., 1987. Signal extraction for finite nonstationary time series. Biometrika 74, 411–421.

Kolmogorov, A.N., 1939. Sur l'interpretation et extrapolation des suites stationnaires. C.R. Acad. Sci. Paris 208, 2043–2045.

Kolmogorov, A.N., 1941. Interpolation and extrapolation von stationären zufälligen Folgen. Bull. Acad. Sci. U.R.S.S. Ser. Math. 5, 3–14.

Koopman, S., 1997. Exact initial Kalman filtering and smoothing for nonstationary time series models. J. Amer. Statist. Assoc. 92, 1630–1638.

Koopman, S., Shepherd, N., Doornik, J., 1999. Statistical algorithms for models in state space using SsfPack 2.2. Econometrics J. 2, 113–166.
Ladiray, D., Quenneville, B., 2001. Seasonal Adjustment with the X-11 Method. Springer, New York.
Pollock, D., 2000. Trend estimation and de-trending via rational square-wave filters. J. Econometrics 99, 317–334.
Pollock, D., 2001. Filters for short non-stationary sequences. J. Forecasting 20, 341–355.
Pollock, D., 2002. Circulant matrices and time-series analysis. Int. J. Math. Educ. Sci. Technol. 33, 213–230.
Sobel, E.L., 1967. Prediction of a noise-distorted, multivariate, non-stationary signal. J. Appl. Probab. 4, 330–342.
Wiener, N., 1949. The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications. Wiley, New York.