RESEARCH REPORT SERIES
(*Statistics #2004-05 Revised 3/3/05)*

An Iterated Parametric Approach to Nonstationary Signal Extraction

Tucker McElroy and Andrew Sutcliffe[1]

Statistical Research Division
U.S. Bureau of the Census
Washington D.C. 20233

Report Issued: September 30, 2004 (Revised 3/3/05)

---

[1]Australian Bureau of Statistics

An Iterated Parametric Approach to Nonstationary Signal Extraction

Tucker McElroy
U.S. Census Bureau
Andrew Sutcliffe
Australian Bureau of Statistics

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 $Y^S = S + I$ and $Y^T = T + I$. This "bootstrapping" signal extraction methodology is reminiscent of X-11's iterated nonparametric approach. The analysis of the iteration scheme provides insight into the algebraic relationship between full model and reduced model signal extraction estimates.

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 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 a very deep 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. Bell considered two main premises – Assumption A and Assumption B – neither of which is typically verifiable from the observed data.

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 B states that the initial values of signal and noise are generated independently of each other, as well as from the differenced signal and noise processes. Although Assumption B is arguably more natural from a modeling standpoint, since it entails that signal and noise be independent for all time, 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, whereas Assumption B does require this unattainable knowledge. 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) and Bell and Hilmer (1991) 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. 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. 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 provides insight into 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.

We note that McElroy and Sutcliffe (2005) is a shortened version of this paper. We 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 Two develops these formulas and the attendant notations and gives a motivating example. Section Three 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 Four, and its rate of convergence is assessed through matrix norms. Section Five discusses the implementation of these ideas, and presents the results of a simulation study and the analysis of U.S. Shoe Store Sales data. We conclude in Section Six, and provide one technical proof in the Appendix B. Appendix A contains some additional material regarding the calculation of finite-sample Mean Squared Errors of signal extraction estimates.

2 Background and Notation

We attempt to follow the notation of Bell and Hilmer (1988), and all formulas are presented in a vector framework. Thus, a sample of n observed data $Y_1, Y_2, \cdots, Y_n$ 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 $Y_1, \cdots, Y_d$ are independent of the differenced trend, differenced seasonal, and the irregular. We refer to this as Assumption A'.

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). Later we will discuss the integrating power series, which 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 represent an over-simplification of 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

$$Y_t^T = T_t + I_t$$
$$Y_t^S = S_t + I_t$$

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:

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

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:

$$\mathbb{E}(T|Y^T) = F_{TI}^T\, Y^T$$
$$\mathbb{E}(S|Y^S) = F_{SI}^S\, Y^S$$

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. 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:

$$\mathbb{E}(T|Y) = F^T_{STI}\,Y$$
$$\mathbb{E}(S|Y) = F^S_{STI}\,Y$$

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. Let Δ^*_S be an $n - d_S$ by n matrix which operates on Y, and let Δ^*_T be an $n - d_T$ by n matrix that also operates on Y with entries given by:

$$(\Delta^*_S)_{ij} = -\delta^S_{i-j+d_S} \qquad (\Delta^*_T)_{ij} = -\delta^T_{i-j+d_T}$$

where $-\delta^S_k$ is the kth coefficient of δ^S. We follow the convention of Bell (1984) in writing these coefficients with a minus sign. As is usual for ARIMA models, $-\delta^S_0 = 1$, and of course the kth coefficient is zero if either k is negative or exceeds d_S. The analogous notation is used for δ^T and the full differencing polynomial δ. More explicitly, we have

$$\Delta^*_S = \begin{bmatrix} -\delta^S_{d_S} & -\delta^S_{d_S-1} & \cdots & 1 & 0 & \cdots & 0 \\ 0 & -\delta^S_{d_S} & \cdots & -\delta^S_1 & 1 & \cdots & 0 \\ \cdots & & & & & & \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Each row of these matrices corresponds to the action of the differencing polynomial on the data; it is easy to see that

$$U^T = \Delta^*_T\,T \quad V^T = \Delta^*_S\,(S+I)$$
$$U^S = \Delta^*_S\,S \quad V^S = \Delta^*_T\,(T+I)$$
$$W^T = \Delta^*_T\,Y^T \quad W^S = \Delta^*_S\,Y^S$$

Now for the next round of differencing, we need matrices that act on $n - d_T$ and $n - d_S$ component vectors, producing $n - d$ values afterwards. So define Δ_S and Δ_T to have the same entries as their cousins Δ^*_S and Δ^*_T, but with dimensions $n - d$ by $n - d_T$ and $n - d$ by $n - d_S$, respectively. Then we have

$$W = \Delta_S U^T + \Delta_T V^T = \Delta_T U^S + \Delta_S V^S$$

Finally, we define Δ to be an $n-d$ by n matrix with entries

$$\Delta_{ij} = -\delta_{i-j+d}$$

Then the following intuitive lemma relates these matrices:

Lemma 1

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

Proof. We prove the first equality:

$$(\Delta_S \Delta_T^*)_{ij} = \sum_{k=1}^{n-d_T} (-\delta^S_{i-k+d_S})(-\delta^T_{k-j+d_T}) = \sum_{l=i+d-n}^{i+d_S-1} (-\delta^S_l)(-\delta^T_{i-j-l+d})$$

The bounds on the sum can be taken from $-\infty$ to ∞, due to $1 \leq i \leq n-d$ and the order of δ^S; but the kth coefficient of δ, by simple algebra, is given by

$$-\delta_k = \sum_l (-\delta^S_l)(-\delta^T_{k-l})$$

Hence the ijth coefficient of the above product of matrices is $-\delta_{i-j+d} = \Delta_{ij}$. $\square$

This provides us with new expressions for W:

$$\begin{aligned} W &= \Delta_T U^S + \Delta_S W^T \\ &= \Delta_S U^T + \Delta_T W^S \\ &= \Delta_T U^S + \Delta_S U^T + \Delta I \end{aligned}$$

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

$$(1-B)^2 U(B) Y_t = (1-\theta B)(1-\Theta B^{12})\epsilon^Y_t \tag{1}$$

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

$$\delta^T(B) = (1-B)^2 \qquad \delta^S(B) = U(B) = 1 + B + \cdots + B^{11} \qquad \delta(B) = (1-B)(1-B^{12})$$

It follows from (1) that

$$W_t = (1-\theta B)(1-\Theta B^{12})\epsilon^Y_t$$

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 (1) into its component parts. If we form the canonical decomposition, assuming admissability (see Hilmer and Tiao (1982)), then we obtain component models

$$U(B)S_t = \theta^S(B)\epsilon_t^S = U_t^S$$
$$(1-B)^2 T_t = \theta^T(B)\epsilon_t^T = U_t^T$$
$$I_t = \epsilon_t^I$$

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

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

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) and Koopman, Shepherd, and Doornik (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. 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 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:*

$$F_{TI}^T = 1_n - \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* = \left(1_n + \Sigma_I \Delta_T^{*\prime} \Sigma_{U^T}^{-1} \Delta_T^*\right)^{-1} \quad (2)$$
$$F_{SI}^S = 1_n - \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* = \left(1_n + \Sigma_I \Delta_S^{*\prime} \Sigma_{U^S}^{-1} \Delta_S^*\right)^{-1}$$

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

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

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

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

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. $\square$

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 (2) 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 (2) are also globally optimal.

In order to express the full model extraction matrices, we utilize notation developed in Bell and Hilmer (1988), which is repeated here for convenient reference. Let $\tilde{\Delta}_T$ be the invertible matrix defined by

$$\tilde{\Delta}_T = \begin{bmatrix} 1_{d_T} & 0 \\ \multicolumn{2}{c}{\Delta_T^*} \end{bmatrix}$$

Then the trend extraction estimate, under Assumption A', is given by

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

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),

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

Let us name this latter matrix $\underline{T}$:

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

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):

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

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,

$$H_1 = \begin{bmatrix} 1_{d_T} \\ {A^T_{d_T+1}}' \\ \cdots \\ {A^T_d}' \end{bmatrix}$$

with the ith entry of the column vector A^T_j given by

$$\sum_{l=0}^{d_T-i} (-\delta^T_l)\xi^T_{j-i-l}$$

for $1 \leq i \leq d_T$. Also ξ^T_k 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

$$H_2 = \begin{bmatrix} 1_{d_S} \\ {A^S_{d_S+1}}' \\ \cdots \\ {A^S_d}' \end{bmatrix}$$

and the ith entry of A^S_j is

$$\sum_{l=0}^{d_S-i} (-\delta^S_l)\xi^S_{j-i-l}$$

for $1 \leq i \leq d_S$. 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

$$C_{1ij} = \xi^T_{i-j-d_T} \qquad C_{2ij} = \xi^S_{i-j-d_S}$$

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

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

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

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

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:*

$$F^T_{STI} = \tilde{\Delta}_T^{-1}\begin{bmatrix} \overline{T} \\ \underline{T} \end{bmatrix} \qquad F^S_{STI} = \tilde{\Delta}_S^{-1}\begin{bmatrix} \overline{S} \\ \underline{S} \end{bmatrix}$$

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 Four, 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

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

Proposition 3 *Define F^S_{SI} and F^T_{TI} as in Proposition 1. We have $0 < \lambda_n(F^S_{SI} F^T_{TI}) \leq \lambda_1(F^S_{SI} F^T_{TI}) < 1$. Also the inverse of $1_n - F^S_{SI} F^T_{TI}$ exists. The series $\sum_{k=0}^{\infty} (F^S_{SI} F^T_{TI})^k$ is convergent with sum equal to $(1_n - F^S_{SI} F^T_{TI})^{-1}$. The same results hold for $F^T_{TI} F^S_{SI}$ in place of $F^S_{SI} F^T_{TI}$.*

Proof. We first show the invertibility of $1_n - F^S_{SI} F^T_{TI}$. Observe that

$$\begin{aligned} 1_n - F^S_{SI} F^T_{TI} &= F^S_{SI} \left((F^S_{SI})^{-1} (F^T_{TI})^{-1} - 1_n \right) F^T_{TI} \\ &= F^S_{SI} \left(\Sigma_I \, \Delta^{*\prime}_S \, \Sigma^{-1}_{U^S} \, \Delta^*_S + \Sigma_I \, \Delta^{*\prime}_T \, \Sigma^{-1}_{U^T} \, \Delta^*_T + \Sigma_I \, \Delta^{*\prime}_S \, \Sigma^{-1}_{U^S} \, \Delta^*_S \, \Sigma_I \, \Delta^{*\prime}_T \, \Sigma^{-1}_{U^T} \, \Delta^*_T \right) F^T_{TI} \end{aligned}$$

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

$$\begin{aligned} &\lambda_n \left(1_n + \Sigma_I \, \Delta^{*\prime}_S \, \Sigma^{-1}_{U^S} \, \Delta^*_S + \Sigma_I \, \Delta^{*\prime}_T \, \Sigma^{-1}_{U^T} \, \Delta^*_T + \Sigma_I \, \Delta^{*\prime}_S \, \Sigma^{-1}_{U^S} \, \Delta^*_S \, \Sigma_I \, \Delta^{*\prime}_T \, \Sigma^{-1}_{U^T} \, \Delta^*_T \right) \\ &= \lambda_n \left((1_n + \Sigma_I^{1/2 \prime} \, \Delta^{*\prime}_S \, \Sigma^{-1}_{U^S} \, \Delta^*_S \, \Sigma_I^{1/2})(1_n + \Sigma_I^{1/2 \prime} \, \Delta^{*\prime}_T \, \Sigma^{-1}_{U^T} \, \Delta^*_T \, \Sigma_I^{1/2}) \right) \\ &= \lambda_n(GH) \end{aligned}$$

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

$$\lambda_n(GH) = \lambda_n(G^{1/2 \prime} H G^{1/2}) = \inf_{x \neq 0} \frac{x' G^{1/2 \prime} H G^{1/2} x}{x' x} = \inf_{y \neq 0} \frac{y' H y}{y' G^{-1} y}$$

We can compute the inverse of G:

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

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

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

for all $x \neq 0$. Hence

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

is nonnegative and bounded. Now this quantity, using the non-negative definiteness of $\Delta_T^{*\prime}\, \Sigma_{U^T}^{-1}\, \Delta_T^*$ and $\Delta_S^{*\prime}\, \Sigma_{W^S}^{-1}\, \Delta_S^*$, 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.*

Proof. 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 $i = 1, \cdots, h$ and $j = 0, 1, \cdots, m_i - 1$; $t = 1, \cdots, n$. 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 $t = 1, 2, \cdots, n$. 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^*. $\square$

Now since $\lambda_n(GH) > 1$, we have

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

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

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

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

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

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

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

Finally, we compute the Schur decomposition of $F_{SI}^S F_{TI}^T$ as follows:

$$Q' F_{SI}^S F_{TI}^T Q = \Lambda + N$$

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 $\theta \geq 0$ and obtain the bound

$$\|(F_{SI}^S F_{TI}^T)^k\|_2 \leq (1+\theta)^{n-1} \big(\lambda_1(F_{SI}^S F_{TI}^T) + \|N\|_F/(1+\theta)\big)^k \tag{3}$$

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

$$\theta = \frac{2\|N\|_F}{1 - \lambda_1(F_{SI}^S F_{TI}^T)} - 1$$

and then

$$\lambda_1(F_{SI}^S F_{TI}^T) + \|N\|_F/(1+\theta) = \frac{1}{2}(1 + \lambda_1(F_{SI}^S F_{TI}^T) < 1$$

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:

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

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

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

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

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 $F^T_{TI} = \Sigma_T\, \Sigma^{-1}_{TI}$, $F^T_{STI} = \Sigma_T\, \Sigma^{-1}_Y$, and $F^S_{STI} = \Sigma_S\, \Sigma^{-1}_Y$, and by orthogonality of components, it follows that $\Sigma_Y = \Sigma_T + \Sigma_S + \Sigma_I$. The proof of the above relation is then

$$\begin{aligned} F^T_{TI} \left(1_n - F^S_{STI}\right) &= \Sigma_T \left(\Sigma_T + \Sigma_I\right)^{-1} \left(1_n - \Sigma_S\, \Sigma^{-1}_Y\right) \\ &= \Sigma_T \left(\Sigma_Y - \Sigma_S\right)^{-1} \left(\Sigma_Y - \Sigma_S\right) \Sigma^{-1}_Y \\ &= \Sigma_T\, \Sigma^{-1}_Y = F^T_{STI} \end{aligned}$$

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

Theorem 1 *Suppose that the assumptions made in the beginning of Section Two are valid, as well as Assumption A'. Let F^T_{TI} and F^S_{SI} be defined by the formula (2). Then the following formulas hold:*

$$\begin{aligned} F^T_{STI} &= F^T_{TI} \left(1_n - F^S_{STI}\right) \\ F^S_{STI} &= F^S_{SI} \left(1_n - F^T_{STI}\right) \end{aligned} \tag{4}$$

These equations can be solved simultaneously to yield

$$\begin{aligned} \left(1_n - F^T_{TI}\, F^S_{SI}\right) F^T_{STI} &= F^T_{TI} \left(1_n - F^S_{SI}\right) \\ \left(1_n - F^S_{SI}\, F^T_{TI}\right) F^S_{STI} &= F^S_{SI} \left(1_n - F^T_{TI}\right) \end{aligned} \tag{5}$$

In addition, the matrices $(1_n - F^S_{SI}\, F^T_{TI})$ and $(1_n - F^T_{TI}\, F^S_{SI})$ are invertible, which results in the explicit formulas

$$\begin{aligned} F^T_{STI} &= \left(1_n - F^T_{TI}\, F^S_{SI}\right)^{-1} F^T_{TI} \left(1_n - F^S_{SI}\right) \\ F^S_{STI} &= \left(1_n - F^S_{SI}\, F^T_{TI}\right)^{-1} F^S_{SI} \left(1_n - F^T_{TI}\right) \end{aligned} \tag{6}$$

Remark 3 Bill Bell has pointed out that this result is true under Assumption B as well, in which case the proof is extremely simple:

$$\begin{aligned} F^T_{STI} Y &= \mathbb{E}[T|Y] = \mathbb{E}[\mathbb{E}[T|Y^T, S]|Y] = \mathbb{E}[(\mathbb{E}[T|Y^T] + \mathbb{E}[T|S])|Y] \\ &= \mathbb{E}[F^T_{TI} Y^T|Y] = F^T_{TI}\mathbb{E}[Y^T|Y] = F^T_{TI}(1 - F^S_{STI})Y \end{aligned}$$

which relies on the orthogonality of Y^T and S in the third equality, and the orthogonality of T and S in the fourth. Also note that we do not make Assumption A for the reduced models Y^T and Y^S, whose filter matrices F^T_{TI} and F^S_{SI} are constructed according to the transformation approach.

Remark 4 The first pair of formulas (4) give an intuitive interpretation of the relationship between F^T_{STI} and F^S_{STI}. 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 (6) 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):

$$F^C_{STI} = 1_n - \Sigma_I \, \Delta' \, \Sigma_W^{-1} \, \Delta \tag{7}$$

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

$$F^C_{STI} = F^T_{STI} + F^S_{STI} \tag{8}$$

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

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

are valid for all Y; hence (8) must hold (a completely algebraic proof of (8) is included in the appendix). From (8) we can write

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

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

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

which uses the fact that

$$\Delta_T^* F_{STI}^T = [0\, 1_{n-d_T}] \tilde{\Delta}_T \tilde{\Delta}_T^{-1} [\overline{T}' \; \underline{T}']' = \underline{T} = \Sigma_{U^T} \, \Delta_S{}' \Sigma_W^{-1} \, \Delta$$

The second equation in (4) has a similar proof.

Now if we solve the pair of equations in (4) we immediately obtain (5). The invertibility of $1 - F_{SI}^S F_{TI}^T$ is demonstrated in Proposition 3. We note in passing that if S and T are stationary, the above formula reduces to

$$(1_n - \Sigma_I \, \Sigma_{Y^S}^{-1}) \, \Sigma_I \, \Sigma_S^{-1} \, \Sigma_Y \, \Sigma_T^{-1} \, (1_n - \Sigma_I \, \Sigma_{Y^T}^{-1}) = \Sigma_S \, \Sigma_{Y^S}^{-1} \, \Sigma_I \, \Sigma_S^{-1} \, \Sigma_Y \, \Sigma_{Y^T}^{-1}$$

whose invertibility is now obvious. With analogous calculations for the second line of (6), the proof is complete. $\square$

For completeness, we next present some other relationships and formulas that immediately follow from the preceding development. The irregular filter is given implicitly in (7) by

$$F_{STI}^I = \Sigma_I \Delta' \Sigma_W^{-1} \Delta \tag{9}$$

The seasonal adjustment filter and detrending filter are denoted by F_{STI}^{TI} and F_{STI}^{SI} respectively; the following corollary develops their properties:

Corollary 1 *Under the same assumptions as Theorem 1*

$$\begin{aligned} F_{STI}^{TI} &= (1 - F_{SI}^S F_{TI}^T)^{-1} (1 - F_{SI}^S) = M_T \Sigma_I \Delta_S^{*\,\prime} \Sigma_{W^S}^{-1} \Delta_S^* \\ F_{STI}^{SI} &= (1 - F_{TI}^T F_{SI}^S)^{-1} (1 - F_{TI}^T) = M_S \Sigma_I \Delta_T^{*\,\prime} \Sigma_{W^T}^{-1} \Delta_T^* \end{aligned} \tag{10}$$

where $M_T = (1_n - F_{SI}^S F_{TI}^T)^{-1}$ *and* $M_S = (1_n - F_{TI}^T F_{SI}^S)^{-1}$. *Also, the following relations hold:*

$$\begin{aligned} F_{TI}^T F_{STI}^{TI} &= F_{STI}^T \\ F_{SI}^S F_{STI}^{SI} &= F_{STI}^S \end{aligned} \tag{11}$$

Proof. By conditional expectations and (4),

$$F_{STI}^{TI} = 1 - F_{STI}^S = F_{TI}^{T\;-1} F_{STI}^T$$

This proves (11), and by using (6) we obtain (10). $\square$

Remark 5 It follows from Corollary 1 that both F_{STI}^T and F_{STI}^{TI} consist of a matrix right-multiplied by Δ_S^*, thereby demonstrating that a seasonal differencing is the first operation to the data.

4 The Main Algorithm and Its Analysis

The idea of the algorithm of this section is to use (4) to define an iteration scheme. The resulting algorithm produces estimates of signal and trend $\hat{S}$ and $\hat{T}$ that satisfy

$$\hat{T} = F_{TI}^{T}(Y - \hat{S})$$
$$\hat{S} = F_{SI}^{S}(Y - \hat{T})$$

which is essentially (4) applied to Y. This is done by essentially by constructing the solutions to the linear system defined by applying (6) 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. When implemented on some airline models – we performed the canonical decomposition into trend, seasonal and irregular in the sense of Hilmer and Tiao (1982) on a Box-Jenkins airline model – the algorithm was essentially convergent by the third iteration. Any initialization of the algorithm can be used.

Theorem 2 *Consider the following algorithm:*

$$\begin{aligned} &\hat{S}^{(0)} \quad is\, any\, given\, vector \\ &for\, i\, =\, 1\, to\, convergence \\ &\quad \hat{T}^{(i)} = F_{TI}^{T}(Y - \hat{S}^{(i-1)}) \\ &\quad \hat{S}^{(i)} = F_{SI}^{S}(Y - \hat{T}^{(i)}) \\ &end\, for \end{aligned}$$

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

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

$$\hat{S}^{(i+1)} = F_{SI}^{S}\,(1_n - F_{TI}^{T})\,Y + F_{SI}^{S}\,F_{TI}^{T}\hat{S}^{(i)}$$

from which it follows that

$$\hat{S}^{(i+1)} - \hat{S}^{(i)} = F_{SI}^{S}\,(1_n - F_{TI}^{T})\,Y - (1_n - F_{SI}^{S}\,F_{TI}^{T})\hat{S}^{(i)}$$

At this stage, note that if $\hat{S}^{(i)} = F_{Y}^{S}\,Y$, then by (5) we have

$$\hat{S}^{(i+1)} - \hat{S}^{(i)} = F_{SI}^{S}\,(1_n - F_{TI}^{T})\,Y - (1_n - F_{SI}^{S}\,F_{TI}^{T})\,F_{STI}^{S}\,Y = 0$$

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

$$\hat{S}^{(i+1)} = \sum_{j=0}^{i} (F_{SI}^{S}\,F_{TI}^{T})^{j}\,F_{SI}^{S}\,(1_n - F_{TI}^{T})\,Y + (F_{SI}^{S}F_{TI}^{T})^{i+1}\hat{S}^{(0)}$$

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

$$\hat{T}^{(i+1)} = \sum_{j=0}^{i} (F_{TI}^{T} F_{SI}^{S})^{j} F_{TI}^{T} (1_n - F_{SI}^{S}) Y + (F_{TI}^{T} F_{SI}^{S})^{i} F_{TI}^{T} (F_{SI}^{S} Y - \hat{S}^{(0)})$$

Now, using the fact that $\sum_{j \geq 0} (F_{TI}^{T} F_{SI}^{S})^{j}$ and $\sum_{j \geq 0} (F_{SI}^{S} F_{TI}^{T})^{j}$ are convergent, as shown in Proposition 3, and the fact that $(F_{TI}^{T} F_{SI}^{S})^{i}$ and $(F_{SI}^{S} F_{TI}^{T})^{i}$ tend to zero as $i \to \infty$, we see that the iterates converge to

$$\hat{S}^{(\infty)} = \sum_{j=0}^{\infty} (F_{SI}^{S} F_{TI}^{T})^{j} F_{SI}^{S} (1_n - F_{TI}^{T}) Y = (1_n - F_{SI}^{S} F_{TI}^{T})^{-1} F_{SI}^{S} (1_n - F_{TI}^{T}) Y = F_{STI}^{S} Y$$
$$\hat{T}^{(\infty)} = \sum_{j=0}^{\infty} (F_{TI}^{T} F_{SI}^{S})^{j} F_{TI}^{T} (1_n - F_{SI}^{S}) Y = (1_n - F_{TI}^{T} F_{SI}^{S})^{-1} F_{TI}^{T} (1_n - F_{SI}^{S}) Y = F_{STI}^{T} Y$$

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 (3). $\square$

Remark 6 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. 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 (1) with parameters $\theta = .6$, $\Theta = .4$, 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, Shepherd, and Doornik (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 .1. Notice that in earlier iterations, some

Figure 1: Seasonal Iterates

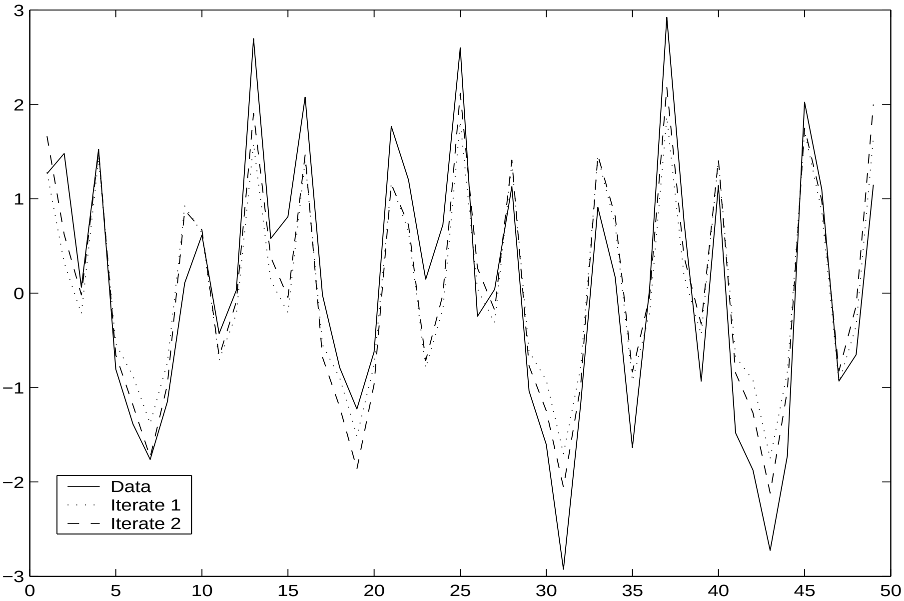


Figure 2: Trend Iterates

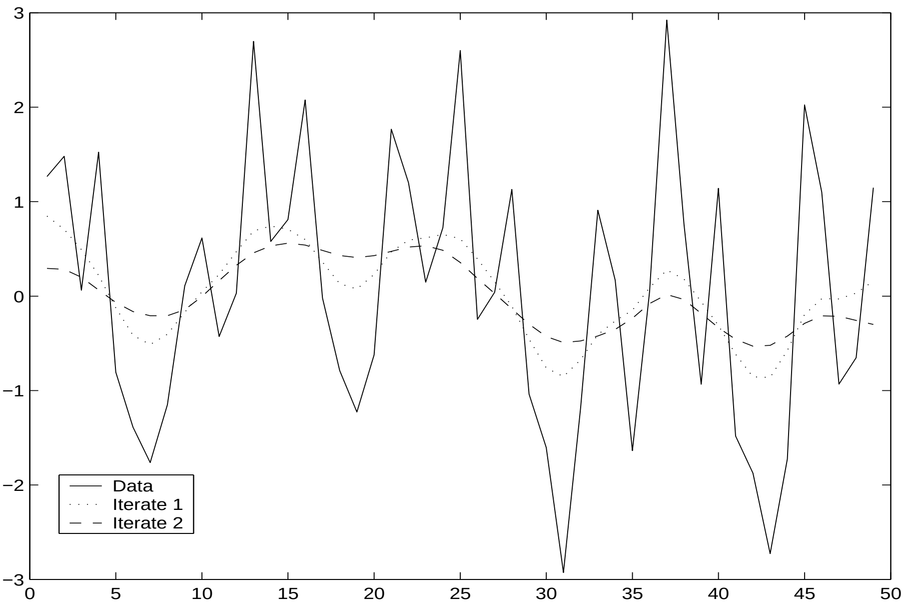
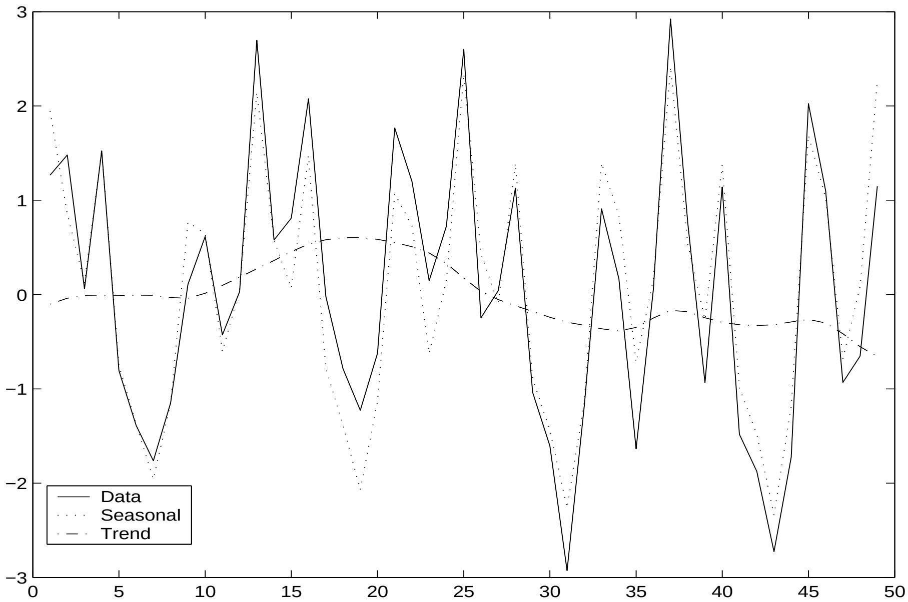


Figure 3: Optimal Estimates

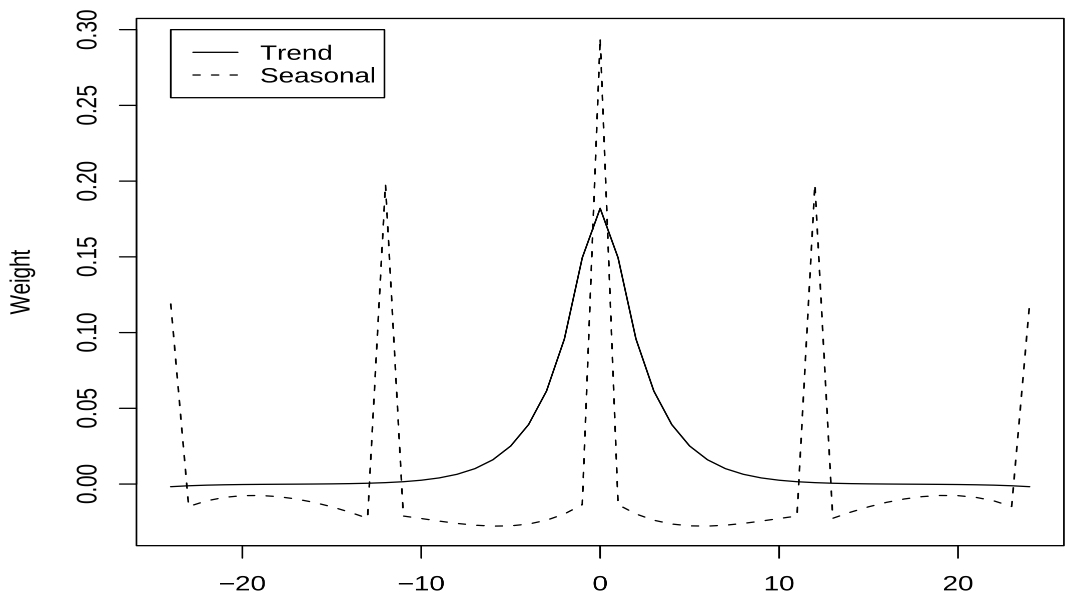


seasonality is present in the trend estimates, and some trend is in the seasonal estimates, but this confusion of signals is gradually weeded out – compare Figures 1 and 2 with Figure 3. In fact, an examination of the filter weights at the center of the sample (Figures 4 and 5) shows that the reduced 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 (4). 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.

Next, we analyzed the U.S. 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 was fit to an airline model as in the simulation study. This produced values of $\theta = .572$ and $\Theta = .336$, and innovation standard deviation .031. The algorithm was initialized with $\hat{S}^{(0)} = 0$ the zero vector, and under the same convergence criterion, it converged in 10 iterations. Figures 6 and 7 show the seasonal and trend iterates for the first five years of data, together with the final estimates. In comparing the squared gains (Figures 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.

Notes on Implementation These plots were produced through Ox code – see Doornik (1998). One may use a State Space Representation and SsfPack (Koopman, Shephard, and Doornik (1999)) to produce the

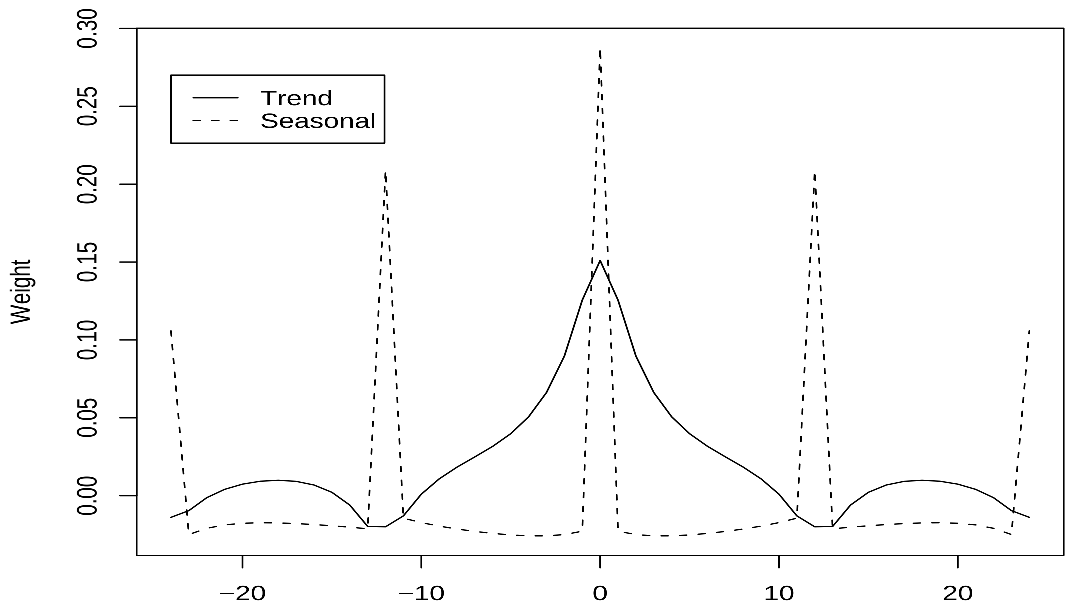
Figure 4: Reduced Model Weights



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 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_{TI}^{T} and F_{SI}^{S} 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

Figure 5: Full Model Weights



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).

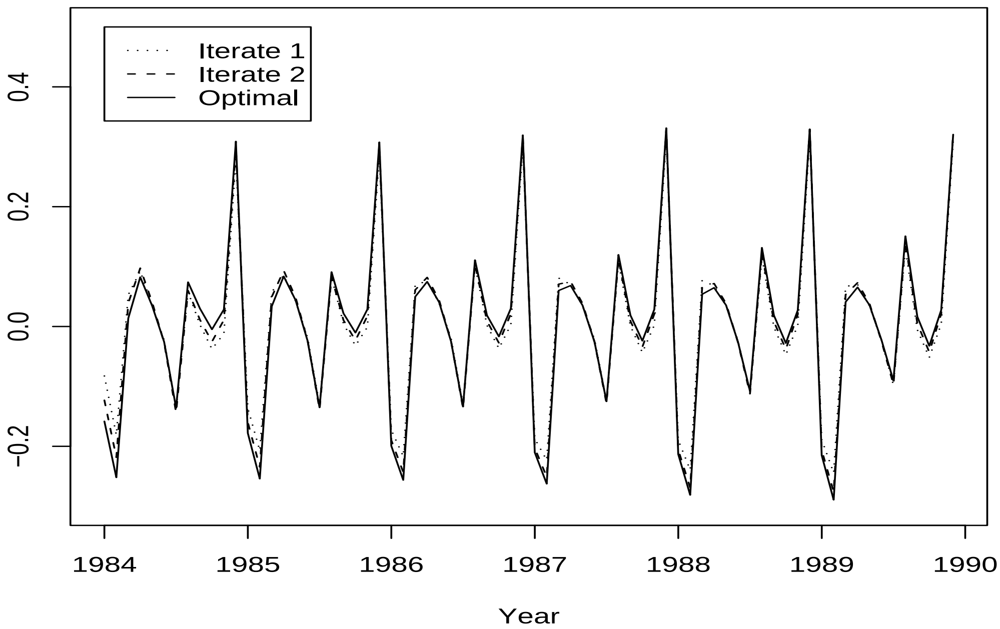
6 Conclusion

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 Four presents a method for building up the correct signal extraction filters from less complicated, more intuitive (and more commonly used) reduced model filters.

Note 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

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

Figure 6: Seasonal Iterates for Shoe Sales



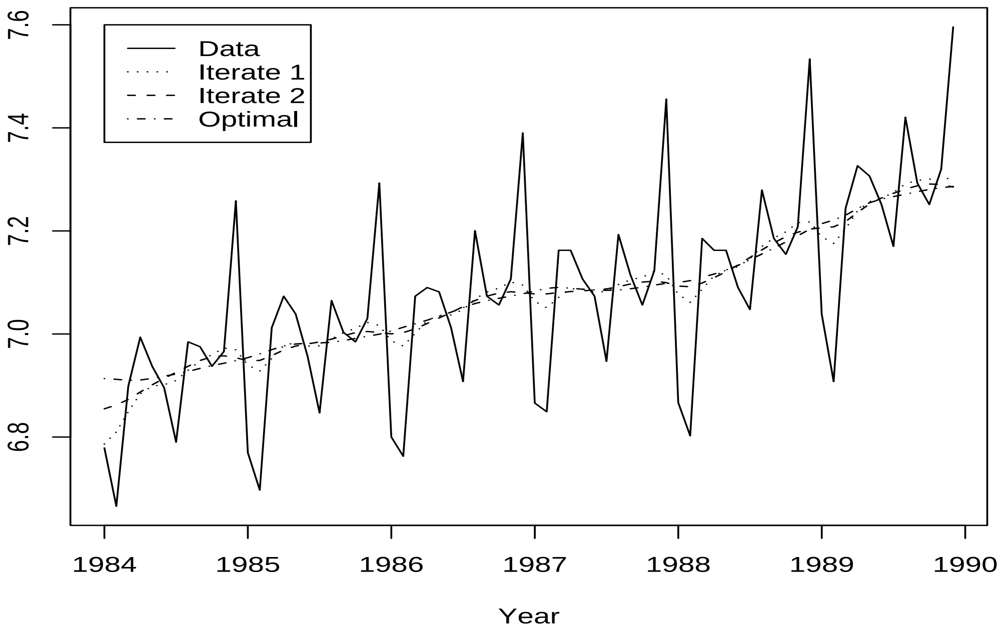
Then this would be followed up with a seasonal extraction filter for a reduced two component model, which is F^S_{SI}. Our resultant seasonal estimate is

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

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 trend and the trend 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 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.

Figure 7: Trend Iterates for Shoe Sales



7 Appendix A: MSE Formulas

In this appendix, we present some formulas for the finite-sample Mean Squared Errors of signal extraction estimates. We introduce the notations

$$M_S = \sum_{j \geq 0} \left(F_{TI}^T F_{SI}^S \right)^j = \left(1 - F_{TI}^T F_{SI}^S\right)^{-1} \qquad M_T = \sum_{j \geq 0} \left(F_{SI}^S F_{TI}^T \right)^j = \left(1 - F_{SI}^S F_{TI}^T\right)^{-1}$$

which are used in the following theorem.

Theorem 3 *The following formulas for the signal extraction MSE's hold:*

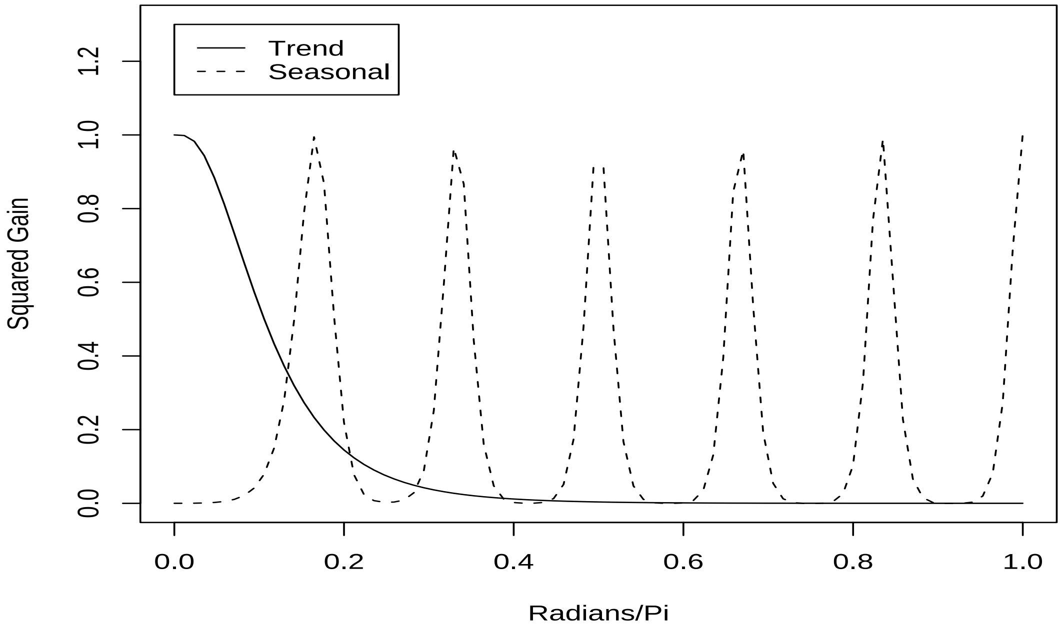
$$MSE[\hat{T}] = Cov[T - \hat{T}] = M_S F_{TI}^T \Sigma_I$$
$$MSE[\hat{S}] = Cov[S - \hat{S}] = M_T F_{SI}^S \Sigma_I$$
$$MSE[\hat{I}] = Cov[I - \hat{I}] = \Sigma_I - \Sigma_I \Delta' \Sigma_W^{-1} \Delta \Sigma_I$$

The MSE for the seasonally adjusted data $\hat{T} + \hat{I}$ is the same as that of the seasonal $\hat{S}$.

Remark 7 These formulas are quite similar to those given for MSE's of signal from a 2-component (i.e., reduced) model, as shown in Bell and Hilmer (1988); it is trivial to show that for the reduced models TI and SI respectively,

$$MSE[\hat{T}] = F_{TI}^T \Sigma_I \qquad MSE[\hat{S}] = F_{SI}^S \Sigma_I$$

Figure 8: Reduced Model Squared Gain



Proof of Theorem 3. We prove the first formula.

$$T - \hat{T} = T - F_{STI}^T Y = (1 - F_{STI}^T)T - F_{STI}^T S - F_{STI}^T I$$

Now since $\Delta_S^* S = U^S$, we have

$$F_{STI}^T S = M_S F_{TI}^T \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} U^S$$

which is stationary. Likewise, using $\Delta_T^* T = U^T$,

$$(1 - F_{STI}^T)T = (F_{SI}^S)^{-1} F_{STI}^S T = (F_{SI}^S)^{-1} M_T F_{SI}^S \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} U^T$$

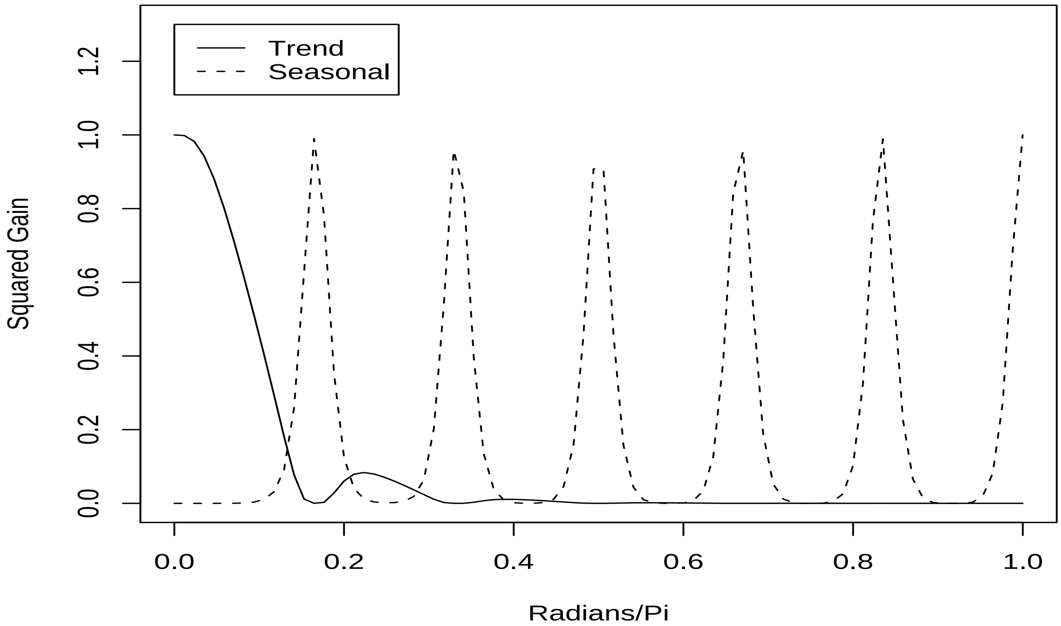
which is also stationary. Now I is orthogonal to U^S which is orthogonal to U^T, so

$$\begin{aligned} Cov[T - \hat{T}] &= M_S F_{TI}^T \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Sigma_{U^S} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I F_{TI}^{T\prime} M_S{}' \\ &+ (F_{SI}^S)^{-1} M_T F_{SI}^S \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Sigma_{U^T} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I F_{SI}^{S\prime} M_T{}' (F_{SI}^S)^{-1\prime} \\ &+ F_{STI}^T \Sigma_I F_{STI}^{T\prime} \end{aligned}$$

Now the third term is easily seen to be

$$M_S F_{TI}^T \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I F_{TI}^{T\prime} M_S{}'$$

Figure 9: Full Model Squared Gain



so that the first and third terms sum to

$$M_S F_{TI}^T \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I F_{TI}^{T\prime} M_S{}'$$

using $\Sigma_{W^S} = \Sigma_{U^S} + \Delta_S^* \Sigma_I \Delta_S^{*\prime}$. The second term, using the above identity again, is:

$$\begin{aligned}
& M_S \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Sigma_{U^T} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I M_S{}' \\
&= M_S \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I M_S{}' - M_S \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I M_S{}' \\
&= M_S F_{TI}^T \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I M_S{}'
\end{aligned}$$

Now adding all the terms together produces

$$\begin{aligned}
Cov[T - \hat{T}] &= M_S F_{TI}^T \left(\Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I + \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I (F_{TI}^T)^{-1\prime} \right) F_{TI}^{T\prime} M_S{}' \\
&= M_S F_{TI}^T \left(\Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I + \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I (1 + \Delta_T^{*\prime} \Sigma_{U^T}^{-1} \Delta_T^* \Sigma_I) \right) F_{TI}^{T\prime} M_S{}' \\
&= M_S F_{TI}^T \Sigma_I \left(\Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Delta_S^* + \Delta_T^{*\prime} \Sigma_{U^T}^{-1} \Delta_T^* \right) \Sigma_I F_{TI}^{T\prime} M_S{}'
\end{aligned}$$

This can be further simplified:

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

implies that

$$Cov[T-\hat{T}] = M_S F_{TI}^T \Sigma_I (1 - F_{SI}^S{}' F_{TI}^T{}') M_S{}' = M_S F_{TI}^T \Sigma_I$$

We know this must be a symmetric matrix, which can be verified also by writing

$$\Sigma_I \left(\Delta_S^*{}' \Sigma_{W^S}^{-1} \Delta_S^* + \Delta_T^*{}' \Sigma_{U^T}^{-1} \Delta_T^* \right) \Sigma_I = \left(F_{TI}^T \right)^{-1} (1 - F_{TI}^T F_{SI}^S) \Sigma_I$$

which yields

$$Cov[T-\hat{T}] = \Sigma_I F_{TI}^T{}' M_S{}'$$

This concludes the calculation of the MSE of $\hat{T}$, and the computation for S is similar. The formula for I is given in Bell and Hilmer (1988). The last statement of the theorem follows from the fact that

$$S + T + I = \hat{S} + \hat{T} + \hat{I}$$

so that

$$\hat{S} - S = (T + I) - (\hat{T} + \hat{I}) \quad \square$$

This theorem can be used to compute MSE's very simply – no Kalman smoother is needed – and the formulas are easy to implement. They also suggest an iterative approach to computing errors, which – as in the case of signal estimation – has the advantage of permitting extreme value adjustment and other non-model based tinkering. The next result computes the variance of the appropriately differenced signals; as can be guessed, the formulas are similar to that for the MSE of the irregular given above.

Proposition 4 *The variance of the differenced seasonal and trend components are given by*

$$Cov[\Delta_T^* \hat{T}] = \Sigma_{U^T} - \Delta_T^* MSE[\hat{T}] \Delta_T^*{}' = \Sigma_{U^T} - \Delta_T^* M_S F_{TI}^T \Sigma_I \Delta_T^*{}'$$
$$Cov[\Delta_S^* \hat{S}] = \Sigma_{U^S} - \Delta_S^* MSE[\hat{S}] \Delta_S^*{}' = \Sigma_{U^S} - \Delta_S^* M_T F_{SI}^S \Sigma_I \Delta_S^*{}'$$

where the formulas for the MSE's are given in Theorem 3.

Proof of Proposition 4. Note that the first equality holds, since the differenced signals have mean zero (being stationary). Now write

$$\Delta_T^* \hat{T} = \Delta_T^* T - \Delta_T^* (T - \hat{T}) \tag{12}$$

The first term is just U^T, whereas the second term is – using the calculations in the proof of Theorem 3 – $\Delta_T^*(1 - F_{STI}^T)T$ plus other terms that are orthogonal to U^T. Hence

$$\begin{aligned} \mathbb{E}[\Delta_T^*(T-\hat{T}) U^T{}'] &= \Delta_T^* \left(F_{SI}^S\right)^{-1} M_T F_{SI}^S \Sigma_I \Delta_T^*{}' \Sigma_{W^T}^{-1} \Sigma_{U^T} \\ &= \Delta_T^* M_S \Sigma_I F_{TI}^T{}' \Delta_T^*{}' \\ &= \Delta_T^* M_S F_{TI}^T \Sigma_I \Delta_T^*{}' \\ &= \Delta_T^* MSE[\hat{T}] \Delta_T^*{}' \end{aligned}$$

using the fact that $F^T_{TI}\Sigma_I = \Sigma_I {F^T_{TI}}'$. Now compute the covariance in equation (12), and we obtain the first formula (with a similar proof for S). $\Box$

Remark 8 Alternative formulas for the MSE of differenced trend and seasonal are due to Bill Bell (personal communication):

$$Cov[\Delta^*_T \hat{T}] = \Sigma_{U^T} \Delta_S{}' \Sigma^{-1}_W \Delta_S \Sigma_{U^T}$$
$$Cov[\Delta^*_S \hat{S}] = \Sigma_{U^S} \Delta_T{}' \Sigma^{-1}_W \Delta_T \Sigma_{U^S}$$

These follow simply from the fact that

$$\Delta^*_T \hat{T} = \Delta^*_T F^T_{STI} Y = \Sigma_{U^T} \Delta_S{}' \Sigma^{-1}_W \Delta Y$$

which has the stated variance. That this expression is equal to the one given in Proposition 4 is not obvious, but can be shown with a little work. Let $B = \Delta^*_T M_S F^T_{TI} \Sigma_I {\Delta^*_T}'$, so that

$$\begin{aligned} B\Sigma^{-1}_W \Delta^*_T &= \Delta^*_T M_S F^T_{TI} (1 - F^T_{TI}) \\ &= \Delta^*_T (1 - F^T_{STI}) F^T_{TI} \\ &= \Delta^*_T F^T_{TI} - \Delta^*_T F^T_{STI} F^T_{TI} \\ &= \Sigma_{U^T} \Sigma^{-1}_{W^T} \Delta^*_T - \Sigma_{U^T} \Delta_S{}' \Sigma^{-1}_W \Delta_S \Sigma_{U^T} \Sigma^{-1}_{W^T} \Delta^*_T \\ &= \left(\Sigma_{U^T} - \Sigma_{U^T} \Delta_S{}' \Sigma^{-1}_W \Delta_S \Sigma_{U^T} \right) \Sigma^{-1}_{W^T} \Delta^*_T \end{aligned}$$

using the fact that $1 - F^T_{STI} = M_S(1 - F^T_{TI})$. Now we can right multiply by

$$\Sigma_I {\Delta^*_T}' \left(\Delta^*_T \Sigma_I {\Delta^*_T}'\right)^{-1} \Sigma_{W^T}$$

(the above inverse exists, since Δ^*_T has rank $n - d_T$) to obtain

$$B = \Sigma_{U^T} - \Sigma_{U^T} \Delta_S{}' \Sigma^{-1}_W \Delta_S \Sigma_{U^T}$$

as desired.

An Algorithm In this section, we present an iterative algorithm to generate the seasonal and trend MSE's. Suppose that J iterations are desired.

Theorem 4 *Consider the following algorithm:*

$$
\begin{aligned}
&\hat{E_T}^{(0)} = 1 \\
&\hat{E_S}^{(0)} = F_{SI}^S \\
&for\, i \,=\, 1\, to\, J \\
&\quad \hat{E_T}^{(i)} = F_{TI}^T \hat{E_S}^{(i-1)} + \hat{E_T}^{(0)} \\
&\quad \hat{E_S}^{(i)} = F_{SI}^S \hat{E_T}^{(i)} \\
&end\, for \\
&\hat{E_T}^{(J)} \leftarrow \hat{E_T}^{(J)} F_{TI}^T \Sigma_I \\
&\hat{E_S}^{(J)} \leftarrow \hat{E_S}^{(J)} \Sigma_I
\end{aligned}
$$

The algorithm converges geometrically fast to $\hat{E_T}^{(\infty)} = M_S F_{TI}^T \Sigma_I$ *and* $\hat{E_S}^{(\infty)} = M_T F_{SI}^S \Sigma_I$.

Proof of Theorem 4. By simple algebra, the Jth iterates are

$$
\begin{aligned}
\hat{E_T}^{(J)} &= \sum_{j=0}^{J} \left(F_{TI}^T F_{SI}^S\right)^j F_{TI}^T \Sigma_I \\
\hat{E_S}^{(J)} &= \sum_{j=0}^{J} \left(F_{SI}^S F_{TI}^T\right)^j F_{SI}^S \Sigma_I
\end{aligned}
$$

Hence, the two-norms of the Jth iterates minus the true MSE's are bounded by

$$
\begin{aligned}
&\sum_{j \geq J+1} \|(F_{TI}^T F_{SI}^S)\|_2^j \|F_{TI}^T \Sigma_I\|_2 \\
&\sum_{j \geq J+1} \|(F_{SI}^S F_{TI}^T)\|_2^j \|F_{SI}^S \Sigma_I\|_2
\end{aligned}
$$

which tend to zero at geometric rate, since the 2-norms are bounded above by one. $\square$

It turns out that the MSE iterates above are approximately equal to the MSE's of the signal extraction iterates:

Proposition 5 *The MSE's of* $\hat{T}^{(J)}$ *and* $\hat{S}^{(J)}$ *are equal to the* J*th iterates in Theorem 4 above, plus some error that is order* $O(\eta^{J+1})$ *for some* $\eta < 1$.

Proof of Proposition 5. Let $M_S^{(J)} = \sum_{j=0}^{J} \left(F_{TI}^T F_{SI}^S\right)^j$, and similarly define $M_T^{(J)}$. Then

$$
T - \hat{T}^{(J)} = (1 - M_S^{(J)} F_{TI}^T (1 - F_{SI}^S)) T - M_S^{(J)} F_{TI}^T (1 - F_{SI}^S) S - M_S^{(J)} F_{TI}^T (1 - F_{SI}^S) I
$$

and the first term simplifies to

$$
1 - M_S^{(J)} F_{TI}^T + M_S^{(J)} F_{TI}^T F_{SI}^S = M_S^{(J+1)} - M_S^{(J)} + M_S^{(J)} (1 - F_{TI}^T)
$$

which operates on T. Now the matrix $M_S^{(J+1)} - M_S^{(J)}$, denoted by $\epsilon^{(J)}$, has 2-norm of order η^{J+1}, where $\eta = \|F_{TI}^T F_{SI}^S\|_2$. It follows that

$$
\begin{aligned}
Cov[T - \hat{T}^{(J)}] &= M_S^{(J)} \Sigma_I \Delta_T^{*\prime} \Sigma_{W^T}^{-1} \Sigma_{U^T} \Sigma_{W^T}^{-1} \Delta_T^* \Sigma_I M_S^{(J)\prime} \\
&+ M_S^{(J)} F_{TI}^T \Sigma_I \Delta_S^{*\prime} \Sigma_{W^S}^{-1} \Sigma_{U^S} \Sigma_{W^S}^{-1} \Delta_S^* \Sigma_I F_{TI}^{T\prime} M_S^{(J)\prime} \\
&+ M_S^{(J)} F_{TI}^T (1 - F_{SI}^S) \Sigma_I (1 - F_{SI}^S)' F_{TI}^{T\prime} M_S^{(J)\prime} \\
&+ O(\eta^{J+1})
\end{aligned}
$$

where the last term is to be interpreted in terms of the matrix 2-norm. Now by the same arguments used in the proof of Theorem 3, the first three terms above simplify to

$$
M_S^{(J)} F_{TI}^T \Sigma_I M_S^{-1\prime} M_S^{(J)\prime} = M_S^{(J)} F_{TI}^T \Sigma_I (1 - (F_{SI}^{S\prime} F_{TI}^{T\prime})^{J+1})
$$

which is $M_S^{(J)} F_{TI}^T \Sigma_I$ plus terms that have matrix 2-norm of order $O(\eta^{J+1})$. As seen in the proof of Theorem 4,

$$
M_S^{(J)} F_{TI}^T \Sigma_I = E_T^{(J)}
$$

as desired. The same proof holds for the seasonal iterates. $\square$

8 Appendix B: Technical Proof

Herein is an algebraic proof of equation (8). We first verify this equation left multiplied by Δ:

$$
\begin{aligned}
\Delta F_{STI}^C &= \Delta - \Delta\, \Sigma_I\, \Delta'\, \Sigma_W^{-1}\, \Delta \\
&= (\Sigma_W - \Delta\, \Sigma_I\, \Delta')\, \Sigma_W^{-1}\, \Delta \\
&= \left(\Delta_T\, \Sigma_{U^S}\, \Delta_T{}' + \Delta_S\, \Sigma_{U^T}\, \Delta_S{}' \right) \Sigma_W^{-1}\, \Delta \\
&= \Delta\, F_{STI}^S + \Delta\, F_{STI}^T
\end{aligned}
$$

where the last line follows from

$$
\Delta\, F_{STI}^T = \Delta_S\, \Delta_T^*\, F_{STI}^T = \Delta_S\, [0\, 1_{n-d_T}] \tilde{\Delta}_T\, F_{STI}^T = \Delta_S\, \underline{T} = \Delta_S\, \Sigma_{U^T}\, \Delta_S{}'\, \Sigma_W^{-1}\, \Delta
$$

and a similar expression for $\Delta\, F_{STI}^S$. Next, we show that (8) is valid when left multiplied by $[1_d\, 0]$. We need to write out $\tilde{\Delta}_T^{-1}$ explicitly; write

$$
\tilde{\Delta}_T = \begin{bmatrix} 1_{d_T} & 0 \\ G_T & H_T \end{bmatrix}
$$

with G_T a $n - d_T$ by d_T matrix, and H_T a square $n - d_T$ dimensional lower triangular matrix. So H_T is invertible, and it is easy to check that

$$
\tilde{\Delta}_T^{-1} = \begin{bmatrix} 1_{d_T} & 0 \\ -H_T^{-1}\, G_T & H_T^{-1} \end{bmatrix}
$$

We use the same notations for $\tilde{\Delta}_S^{-1}$. Then it follows that

$$
\begin{aligned}
[1_d\, 0]\, F_{STI}^T &= \begin{bmatrix} \overline{T} \\ -[1_{d_S}\, 0] H_T^{-1}\, G_T\, \overline{T} + [1_{d_S}\, 0] H_T^{-1} \underline{T} \end{bmatrix} \\
&= \left(\begin{bmatrix} 1_{d_T} \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1_{d_S} \end{bmatrix} [1_{d_S}\, 0]\, H_T^{-1}\, G_T \right) \overline{T} + \begin{bmatrix} 0 \\ 1_{d_S} \end{bmatrix} [1_{d_S}\, 0] H_T^{-1} \underline{T}
\end{aligned}
$$

In a similar fashion, we compute

$$
[1_d\, 0]\, F_{STI}^S = \left(\begin{bmatrix} 1_{d_S} \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 1_{d_T} \end{bmatrix} [1_{d_T}\, 0]\, H_S^{-1}\, G_S \right) \overline{S} + \begin{bmatrix} 0 \\ 1_{d_T} \end{bmatrix} [1_{d_T}\, 0] H_S^{-1} \underline{S}
$$

Now, the actual entries of H_T and H_S are

$$
(H_T)_{ij} = -\delta_{i-j}^T \qquad (H_S)_{ij} = -\delta_{i-j}^S
$$

from which the inverses are easily calculated to be

$$
(H_T)_{ij}^{-1} = \xi_{i-j}^T \qquad (H_S)_{ij}^{-1} = \xi_{i-j}^S
$$

Likewise we can write

$$
G_T = \Delta_T^* \begin{bmatrix} 1_{d_T} \\ 0 \end{bmatrix} \qquad G_S = \Delta_S^* \begin{bmatrix} 1_{d_S} \\ 0 \end{bmatrix}
$$

Now $[0\, 1_{d_S}]'[1_{d_S}\, 0]$ times H_T^{-1} is just C_1 times $[1_{d_S}\, 0_{d_S\, n-d}]$, with a similar result involving C_2 for F_{STI}^S. Next, we claim that the coefficient matrices of $\overline{T}$ and $\overline{S}$ are

$$
\begin{aligned}
[1_{d_T}\, 0]' - C_1\, [1_{d_S}\, 0]\, \Delta_T^* [1_{d_T}\, 0]' &= H_1 \\
[1_{d_S}\, 0]' - C_2\, [1_{d_T}\, 0]\, \Delta_S^* [1_{d_S}\, 0]' &= H_2
\end{aligned} \tag{13}
$$

First, computation reveals that

$$
\begin{aligned}
(C_2\, [1_{d_T}\, 0]\, \Delta_S^*)_{ij} &= \sum_{k=1}^{n-d_S} \xi_{i-k-d_S}^S (-\delta_{k-j+d_S}^S) \\
&= \sum_{l=1-j+d_S}^{d_S} -\delta_l^S \xi_{i-j-l}^S \\
&= \sum_{l=0}^{d_S-j} \delta_l^S \xi_{i-j-l}^S + 1_{\{i=j\}}
\end{aligned}
$$

which is the jth entry of the column vector $-A_i^S$, plus the Kronecker delta $1_{\{i=j\}}$. This calculation uses a change of variable, the fact that δ^S is an order d_S polynomial, and the coefficient expansion of $\delta^S(x)\xi^S(x) = 1$, as well as the definition of A_i^S. Representing this result in matrix form gives

$$
C_2\, [1_{d_T}\, 0]\, \Delta_S^* = \begin{bmatrix} 0 & 0 \\ -A_{d_S+1}^{S'} & 1_{d_T} \\ \cdots & \end{bmatrix} [1_d\, 0]
$$

From this it follows that

$$C_2\,[1_{d_T}\,0]\,\Delta_S^* = [1_d\,0] - H_2\,[1_{d_T}\,0] \tag{14}$$
$$C_1\,[1_{d_S}\,0]\,\Delta_T^* = [1_d\,0] - H_1\,[1_{d_S}\,0]$$

A little matrix algebra now produces (13). Hence

$$[1_d\,0]\,F_{STI}^T = H_1\overline{T} + C_1\,[1_{d_T}\,0]\,\Sigma_{U^T}\,\Delta_S{}'\,\Sigma_W^{-1}\,\Delta$$
$$[1_d\,0]\,F_{STI}^S = H_2\overline{S} + C_2\,[1_{d_S}\,0]\,\Sigma_{U^S}\,\Delta_T{}'\,\Sigma_W^{-1}\,\Delta$$

Now we must investigate the initial value estimation matrices $\overline{T}$ and $\overline{S}$. Define

$$Q = [1_d\,0] - \Big(C_1\,[1_{d_T}\,0]\,\Sigma_{U^T}\,\Delta_S{}'\,\Sigma_W^{-1}\,\Delta + C_2\,[1_{d_S}\,0]\,\Sigma_{U^S}\,\Delta_T{}'\,\Sigma_W^{-1}\,\Delta\Big)$$

Then, using the fact that

$$\Sigma_{W^S} = \Sigma_{U^S} + \Delta_S^*\,\Sigma_I\,\Delta_S^{*\,'}$$
$$\Sigma_{W^T} = \Sigma_{U^T} + \Delta_T^*\,\Sigma_I\,\Delta_T^{*\,'}$$

we can write

$$\overline{T} = [1_{d_T}\,0]\,J\left(Q - C_2\,[1_{d_S}\,0]\,\Delta_S^*\,(1 - F_{STI}^C)\right)$$
$$\overline{S} = [0\,1_{d_S}]\,J\left(Q - C_1\,[1_{d_T}\,0]\,\Delta_T^*\,(1 - F_{STI}^C)\right)$$

Putting this all together, using $[H_1\,0]J + [0\,H_2]J = 1_d$, we obtain

$$\begin{aligned}
[1_d\,0]\,F_{STI}^T + [1_d\,0]\,F_{STI}^S &= H_1\,\overline{T} + H_2\,\overline{S} + [1_d\,0] - Q \\
&= [H_1\,0]\,JQ + [0\,H_2]\,JQ + [1_d\,0] - Q \\
&\quad - ([H_1\,0]\,JC_2\,[1_{d_S}\,0]\,\Delta_S^* + [0\,H_2]\,JC_1\,[1_{d_T}\,0]\,\Delta_T^*)\,(1 - F_{STI}^C) \\
&= [1_d\,0] - ([H_1\,0]\,JC_2\,[1_{d_S}\,0]\,\Delta_S^* + [0\,H_2]\,JC_1\,[1_{d_T}\,0]\,\Delta_T^*)\,(1 - F_{STI}^C)
\end{aligned}$$

Using the above (14), we have the coefficient of $1 - F_{STI}^C$ equal to

$$\begin{aligned}
&[H_1\,0]\,J\,([1_d\,0] - [H_2\,0]) + [0\,H_2]\,J\,([1_d\,0] - [H_1\,0]) \\
&= [1_d\,0] - J^{-1}\begin{bmatrix} 1_{d_T} & 0 \\ 0 & 0 \end{bmatrix} JJ^{-1}\begin{bmatrix} 0 & 0 \\ 1_{d_S} & 0 \end{bmatrix} - J^{-1}\begin{bmatrix} 0 & 0 \\ 0 & 1_{d_S} \end{bmatrix} JJ^{-1}\begin{bmatrix} 1_{d_T} & 0 \\ 0 & 0 \end{bmatrix} \\
&= [1_d\,0]
\end{aligned}$$

Thus we've shown (8) left multiplied by $[1_d\,0]$, and hence

$$\tilde{\Delta}F_{STI}^T + \tilde{\Delta}F_{STI}^S = \tilde{\Delta}F_{STI}^C$$

where

$$\tilde{\Delta} = \begin{bmatrix} 1_d & 0 \\ & \Delta \end{bmatrix}$$

is an invertible matrix. Inverting this expression now yields (8) as desired.

References

[1] Anderson, B. and Moore, J. (1979). *Optimal Filtering.* Prentice Hall, Englewood Cliffs, NJ.

[2] Ansley, C. and Kohn, R. (1985). Estimation, Filtering, and Smoothing in State Space Models with Incompletely Specified Initial Conditions. *The Annals of Statistics* **13**, 1286 – 1316.

[3] Axelsson, O. (1996). *Iterative Solution Methods.* Cambridge University Press, Cambridge.

[4] Bell, W. (1984). Signal Extraction for Nonstationary Time Series. *The Annals of Statistics* **12**, 646 –664.

[5] Bell, W. (2004). "On RegComponent Time Series Models and Their Applications," in *State Space and Unobserved Component Models: Theory and Applications,* eds. Andrew C. Harvey, Siem Jan Koopman, and Neil Shephard, Cambridge, UK: Cambridge University Press, forthcoming.

[6] Bell, W. and 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.

[7] Bell, W. and Hilmer, S. (1991). Initializing the Kalman Filter for Nonstationary Time Series Models. *Journal of Time Series Analysis* **12**, 283-300.

[8] Cleveland, W. and Tiao, G. (1976). Decomposition of Seasonal Time Series: A Model for the Census X-11 Program. *Journal of the American Statistical Association* **71**, 581 – 587.

[9] Doornik, J. (1998). *Object-oriented Matrix Programming using Ox 2.0.* London: Timberlake Consultants Press.

[10] Golub, G. and Van Loan, C. (1996). *Matrix Computations.* The Johns Hopkins University Press, Baltimore and London.

[11] Hannan, Edward J. (1967). Measurement of a Wandering Signal Amid Noise. *Journal of Applied Probability* **4**, 90-102.

[12] Harvey, A. and Trimbur, T. (2003). General Model-Based Filters for Extracting Cycles and Trends in Economic Time Series. *The Review of Economics and Statistics* **85** (2), 244 – 255.

[13] Henrici, P. (1988). *Applied and Computational Complex Analysis.* Volume 1. John Wiley and Sons, New York.

[14] Hilmer, S. and Tiao, G. (1982). An ARIMA-Model-Based Approach to Seasonal Adjustment. *Journal of the American Statistical Association* **77**, 377, 63 – 70.

[15] Hodrick, R.J. and Prescott, E.C. (1997). Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking* **24**, 1–16.

[16] Kohn, R. and Ansley, C.F. (1986). Estimation, Prediction, and Interpolation for ARIMA Models with Missing Data. *Journal of the American Statistical Association* **81**, 751 – 761.

[17] Kohn, R. and Ansley, C.F. (1987). Signal Extraction for Finite Nonstationary Time Series. *Biometrika* **74**, 411 – 421.

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

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

[20] Koopman, S., Shepherd, N., and Doornik, J. (1999). Statistical Algorithms for Models in State Space Using SsfPack 2.2. *Econometrics Journal* **2**, 113-166.

[21] McElroy, T., and Sutcliffe, A. (2005). An Iterated Parametric Approach to Nonstationary Signal Extraction. To appear in *Computational Statistics and Data Analysis*.

[22] Pollock, D. (2000). Trend estimation and de-trending via rational square-wave filters. *Journal of Econometrics* **99**, 317–334.

[23] Pollock, D. (2001). Filters for Short Non-Stationary Sequences. *Journal of Forecasting* **20**, 341– 355.

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

[25] Wiener, N. (1949). *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series With Engineering Applications.* Wiley, New York.