Econometric Theory, **24**, 2008, 988–1009. Printed in the United States of America.
doi:10.1017/S0266466608080389

MATRIX FORMULAS FOR NONSTATIONARY ARIMA SIGNAL EXTRACTION

TUCKER MCELROY
U.S. Census Bureau

The paper provides general matrix formulas for minimum mean squared error signal extraction for a finitely sampled time series whose signal and noise components are nonstationary autoregressive integrated moving average processes. These formulas are quite practical; in addition to being simple to implement on a computer, they make it possible to easily derive important general properties of the signal extraction filters. We also extend these formulas to estimates of future values of the unobserved signal, and we show how this result combines signal extraction and forecasting.

1. INTRODUCTION

We consider signal extraction for finitely sampled nonstationary time series data that can be transformed to a mean zero, covariance stationary process by differencing operators. We suppose that the signal and noise components are nonstationary and that differencing operators exist that transform the signal and noise components into weakly stationary time series. Signal extraction for infinitely large samples of such series has a long history, including Hannan (1967), Sobel (1967), Cleveland and Tiao (1976), Bell (1984), and Bell and Martin (2004). The unpublished report Bell and Hillmer (1988)—whose basic results are summarized in Bell (2004)—treats the finite-sample case, presenting matrix formulas for the mean square optimal time-varying filters. One drawback of Bell and Hillmer's approach is its need for the separate estimation of initial values of nonstationary signal and noise components, resulting in formulas that are awkward to implement. Pollock (2001) also relies upon estimation of these initial values (and assumes that the noise component is stationary). McElroy and Sutcliffe (2006) furnishes an improvement over the Bell and Hillmer (1988) formulas but only for a specific type of unobserved components model. Assuming that the signal extraction problem is formulated as an autoregressive integrated moving average (ARIMA) signal plus ARIMA noise (we later describe how a multiple unobserved ARIMA components model can be recast into a

The author thanks David Findley for careful reading of this paper, which led to substantial improvements. I also thank the referees for suggesting the "square root information" approach to computing F and M and also for many other improvements. This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the author and not necessarily those of the U.S. Census Bureau. Address correspondence to Tucker McElroy, Statistical Research Division, U.S. Census Bureau, 4700 Silver Hill Road, Washington, DC 20233-9100; e-mail: mcelroy@census.gov.

 © 2008 Cambridge University Press 0266-4666/08 $15.00

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

two-component model), we show in this paper an especially simple formula that does not involve the estimation of initial values. This novel result is the main content of the paper at hand. This and other formulas discussed here are quite practical in that they can be used to derive important general properties of the filters, while also being simple to implement on a computer.

In the context of finite-sample model-based signal extraction, one popular approach—that utilized in SEATS (see the TSW Manual by Maravall and Caporello, 2004, available at www.bde.es)—has been to apply the bi-infinite filters of Bell (1984) (which are the generalizations of the Wiener–Kolmogorov signal extraction filters to the nonstationary case) to the finite sample extended with infinitely many forecasts and backcasts. The result can be calculated exactly with the aid of an algorithm of Tunnicliffe-Wilson (see Burman, 1980). Although easy to implement, this method does not readily reveal properties of finite-sample filters. Moreover, it cannot produce correct finite-sample mean squared errors (MSEs) for the signal estimates. Another approach is to construct the signal plus noise model in state space form (Durbin and Koopman, 2001) and apply the appropriate state space smoother (Kailath, Sayed, and Hassibi, 2000) to extract the signal. Efficient algorithms exist to obtain the time-varying signal extraction filters from the state space smoother, if desired (Koopman and Harvey, 2003); of course, these methods provide recursive formulas rather than explicit algebraic formulas. Thus, the state space approach cannot reveal certain fundamental properties of the signal extraction filters that are obvious from the matrix formulas; see Sections 4.1 and 4.2. In addition, the matrix approach readily provides the full covariance matrix of the signal error, a quantity that is useful in certain applications; see Findley, McElroy, and Wills (2004). Hence, there is both need and appeal for having explicit, readily implemented matrix formulas for nonstationary signal extraction. Note also that one of the original motivations for the matrix approach to signal extraction of Bell and Hillmer (1988) was to provide a method of signal extraction for models that could not be put into state space form, for example, long memory models. Although our results are presented in the ARIMA-model-based framework, generalizations to long memory and heteroskedastic models are discussed also.

We present background material on signal extraction, as well as the main theoretical results, in Section 2. Section 3 extends these results to estimating an unobserved signal at future times. Some applications of the matrix formulas are provided in Section 4. Examples of optimal finite-sample seasonal adjustment and trend filters, along with their gain functions, are discussed in Section 5. Derivations of the basic formulas are contained in the Appendix.

2. MATRIX FORMULAS

Consider a nonstationary time series Y_t that can be written as the sum of two possibly nonstationary components S_t and N_t, the signal and the noise:

$$Y_t = S_t + N_t. \tag{1}$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

Following Bell (1984), we let Y_t be an integrated process such that $W_t = \delta(B)Y_t$ is weakly stationary. Here B is the backshift operator, and $\delta(z)$ is a polynomial with all roots located on the unit circle of the complex plane (also, $\delta(0) = 1$ by convention). This $\delta(z)$ is referred to as the differencing operator of the series, and we assume that it can be factored into relatively prime polynomials $\delta^S(z)$ and $\delta^N(z)$ (i.e., polynomials with no common zeros), such that the series

$$U_t = \delta^S(B)S_t \qquad V_t = \delta^N(B)N_t \tag{2}$$

are mean zero weakly stationary time series, which are uncorrelated with one another. Note that $\delta^S = 1$ and/or $\delta^N = 1$ are included as special cases. (In these cases either the signal or the noise or both are stationary.) We let d be the order of δ, and d_S and d_N are the orders of δ^S and δ^N; because the latter operators are relatively prime, $\delta = \delta^S \cdot \delta^N$ and $d = d_S + d_N$.

There are many examples from econometrics and engineering that conform to this scheme. In the context of component estimation for seasonal time series, for example, the data typically consist of seasonal, trend, and irregular components (Gómez and Maravall, 2001):

$$Y_t = \mathcal{S}_t + \mathcal{T}_t + \mathcal{I}_t. \tag{3}$$

Alternatively, a cycle component (Durbin and Koopman, 2001) is included also:

$$Y_t = \mathcal{S}_t + \mathcal{T}_t + \mathcal{C}_t + \mathcal{I}_t. \tag{4}$$

Now although there are three or four unobserved components, we can always rewrite the models in terms of two components (signal and noise). For example, if we are interested in seasonally adjusting the series, we identify the seasonal component with the noise, and the sum of the remaining components becomes the signal of interest. That is, $N_t = \mathcal{S}_t$ and $S_t = \mathcal{T}_t + \mathcal{I}_t$ or $S_t = \mathcal{T}_t + \mathcal{C}_t + \mathcal{I}_t$, depending on whether (3) or (4) holds, respectively. Typically, the seasonal component is nonstationary with $\delta^N(z) = 1 + z + z^2 + \cdots + z^{11}$ for monthly data, and the trend is nonstationary too. For a twice-integrated trend (and assuming that the irregular and cyclical components are stationary, which is usually the case) $\delta^S(z) = (1 - z)^2$. If instead we assume (4) (with the same ARIMA component models) and are interested in estimating the business cycle, then $S_t = \mathcal{C}_t$ and $N_t = \mathcal{S}_t + \mathcal{T}_t + \mathcal{I}_t$; in this case $\delta_S(z) = 1$ and $\delta_N(z) = (1 - z)(1 - z^{12})$. In this fashion, any number of multiple unobserved components can be reduced to two components, where the signal is defined as the sum of those components that are of interest and the noise consists of the rest. We will see in Section 5.2 that one can apply the matrix formulas for signal extraction without having to derive new ARIMA models for each separate combination of components. The matrix approach shares this property with the Wiener–Kolmogorov approach of SEATS, which only requires a partitioning of the autocovariance generating function for the observed series.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

As in Bell and Hillmer (1988), we assume that Assumption A of Bell (1984) holds for the component decomposition, and we treat the case of a finite sample with $t = 1, 2, \ldots, n$ for $n > d$. Assumption A states that the initial d values of Y_t, that is, the variables $Y_* = (Y_1, Y_2, \ldots, Y_d)$, are independent of $\{U_t\}$ and $\{V_t\}$. For a discussion of the implications of this assumption, see Bell (1984) and Bell and Hillmer (1988); in particular Bell (1984) discusses how the initial values for both the signal and noise components may be generated such that Assumption A holds. Note that mean square optimal signal extraction filters derived under Assumption A agree exactly with the filters implicitly used by a properly initialized state space smoother; see Bell and Hillmer (1991, 1992). A further assumption made is that $\{U_t\}$ and $\{V_t\}$ are uncorrelated with one another.

Now we can write (2) in a matrix form, as follows. Let Δ be an $(n - d) \times n$ matrix with entries given by $\Delta_{ij} = \delta_{i-j+d}$ (with the convention that $\delta_k = 0$ if $k < 0$ or $k > d$). The analogously defined matrices Δ_S and Δ_N have entries given by the coefficients of $\delta^S(z)$ and $\delta^N(z)$ but are $(n - d_S) \times n$- and $(n - d_N) \times n$-dimensional, respectively. This means that each row of these matrices consists of the coefficients of the corresponding differencing polynomial, horizontally shifted in an appropriate fashion. Hence

$$W = \Delta Y \qquad U = \Delta_S S \qquad V = \Delta_N N,$$

where Y is the transpose of $(Y_1, Y_2, \ldots, Y_n)$ and W, U, V, S, and N are analogously defined. We will denote the transpose of Y by Y'. To express

$$W_t = \delta^N(B) U_t + \delta^S(B) V_t \tag{5}$$

in matrix form we need to define further differencing matrices $\underline{\Delta}_N$ and $\underline{\Delta}_S$ with row entries $\delta^N_{i-j+d_N}$ and $\delta^S_{i-j+d_S}$ given by the coefficients of $\delta^N(z)$ and $\delta^S(z)$, respectively, which are $(n - d) \times (n - d_S)$- and $(n - d) \times (n - d_N)$-dimensional. It follows from Lemma 1 of McElroy and Sutcliffe (2006) that

$$\Delta = \underline{\Delta}_N \Delta_S = \underline{\Delta}_S \Delta_N. \tag{6}$$

Then we can write down the matrix version of (5):

$$W = \underline{\Delta}_N U + \underline{\Delta}_S V. \tag{7}$$

For each $1 \leq t \leq n$, the minimum MSE signal extraction estimate is $\hat{S}_t = \mathbb{E}[S_t | Y]$. This can be expressed as a certain linear function of the data vector Y when the data are Gaussian. This estimate is also the minimum MSE *linear* estimate when the data are non-Gaussian. For the remainder of the paper, we do not assume Gaussianity, and by optimality we always refer to the minimum MSE *linear* estimate. Writing $\hat{S} = (\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n)'$, the coefficients of these linear functions form the rows of a matrix F:

$$\hat{S} = FY.$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

The last row of F, for example, corresponds to the concurrent filter, that is, the one-sided filter used to extract a signal at "time present." Because of symmetry properties of F (discussed in Sect. 4.1), this concurrent filter is the reverse of the first row of F.

For any random vector X, let Σ_X denote its covariance matrix. If the process $\{X_t\}$ is weakly stationary, then the covariance matrix Σ_X is symmetric Toeplitz, that is, the ijth entry only depends on $i - j$. With this notation, we can now state the signal extraction formulas.

THEOREM 1. *Suppose that Assumption A holds for the model decomposition (1), that δ^N and δ^S share no common zeros, and that $\{U_t\}$ and $\{V_t\}$ are mean zero weakly stationary, uncorrelated with one another, and purely nondeterministic. Then the minimum MSE linear estimate of S is given by $\hat{S} = FY$, where*

$$F = (\Delta_S' \Sigma_U^{-1} \Delta_S + \Delta_N' \Sigma_V^{-1} \Delta_N)^{-1} \Delta_N' \Sigma_V^{-1} \Delta_N. \tag{8}$$

The covariance matrix of $\hat{S} - S$ is given by M^{-1}, where

$$M = \Delta_S' \Sigma_U^{-1} \Delta_S + \Delta_N' \Sigma_V^{-1} \Delta_N. \tag{9}$$

It also follows from our assumptions that all matrix inverses exist.

Remark 1. When the signal is stationary these formulas reduce to (so $\Delta_N = \underline{\Delta}_N = \Delta$)

$$F = \Sigma_S \Delta' \Sigma_W^{-1} \Delta,$$

$$M^{-1} = \Sigma_S - \Sigma_S \Delta' \Sigma_W^{-1} \Delta \Sigma_S$$

(cf. Bell and Hillmer, 1988; Bell, 2004). Likewise, when the noise is stationary we have

$$F = 1 - \Sigma_N \Delta' \Sigma_W^{-1} \Delta,$$

$$M^{-1} = \Sigma_N - \Sigma_N \Delta' \Sigma_W^{-1} \Delta \Sigma_N,$$

and $\Delta_S = \underline{\Delta}_S = \Delta$. These formulas are proved by simply manipulating (8) and (9), using the fact that

$$(1 + \Sigma_S \Delta' \Sigma_V^{-1} \Delta)^{-1} = 1 - \Sigma_S \Delta' \Sigma_W^{-1} \Delta.$$

Remark 2. The assumption that $\{U_t\}$ and $\{V_t\}$ are weakly stationary in Theorem 1 can be relaxed somewhat. To implement (8) and (9) it is not necessary that Σ_U and Σ_V be Toeplitz, only invertible. For example, Nguyen, Bell, and Gomish (2002) discuss models for a sampling error component with variances that evolve over time, yielding a non-Toeplitz covariance matrix. In particular, suppose that the noise is a sampling error component given by $N_t = h_t e_t$, where

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

h_t is a positive deterministic quantity that generates a heteroskedastic effect and e_t is a stationary time series, for example, an AR(2). Then we can write $\Sigma_N = D(h)\Sigma_e D(h)$, where $D(h)$ denotes a diagonal matrix with entries given by $h_1, h_2, \ldots, h_n$ and Σ_e is the Toeplitz covariance matrix of $\{e_t\}$. Then Σ_N is invertible, with inverse given by

$$\Sigma_N^{-1} = D(h)^{-1}\Sigma_e^{-1}D(h)^{-1}.$$

Bell (2004) also discusses seasonal heteroskedasticity models, where the differenced seasonal component has a non-Toeplitz covariance matrix. Theorem 1 still applies in these sorts of situations.

Because the matrices in M are roughly of dimension equal to the sample size, their inversion will not be overly burdensome. More precisely, the computation of Σ_U^{-1} involves the inversion of a $(n - d_S)$-dimensional matrix; when $\{U_t\}$ is stationary, the Toeplitz structure of Σ_U can be utilized so that its inversion requires order n^2 operations—see Golub and Van Loan (1996). Then to find a particular entry of M^{-1}, we can use the LU decomposition (Golub and Van Loan, 1996, p. 121), obtaining the result in $O(n^3)$ operations. Hence, the error covariance matrix M^{-1} can be computed fairly quickly, and the procedure is easy to implement on a computer (see the discussion in Sect. 5.1). In contrast, a state space approach will usually be faster (as will be the implementation of the Wiener–Kolmogorov filter in SEATS), which is important for large sample sizes (for details, see de Jong and MacKinnon, 1988; Durbin and Quenneville, 1997). The main advantage in the matrix approach lies in the ease of implementation.

Note that if we have any new "signal" defined by HS, where H is a $1 \times n$-dimensional matrix, then it will be estimated by the matrix HF and will have error covariance matrix $HM^{-1}H'$. This calculation will potentially involve all the entries of M^{-1}, not just the near-diagonal entries that state space smoothing algorithms are capable of producing. Applications that require a full knowledge of M^{-1} are benchmarking (Durbin and Quenneville, 1997), revision statistics, growth rates, and signal extraction diagnostics (Findley et al., 2004).

3. COMPONENT FORECASTS

We now present an extension of these matrix formulas to the forecasting of an unobserved component. More precisely, suppose that it is desired to estimate some future values of the signal, denoted by $S_f = (S_{n+1}, \ldots, S_{n+h})'$ (the f subscript denotes the "future"), but as before only the data $Y = (Y_1, \ldots, Y_n)'$ are available. As a related problem, we first consider the linear projection of S_f onto $S = (S_1, \ldots, S_n)'$, following Bell (2004). The optimal (as always, this signifies the minimum MSE among linear estimates) estimate is given by

$$\hat{S}_f = DS. \tag{10}$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

Here D is the $h \times n$ matrix defined by (11), which follows. We arrive at the formula for D through the following discussion, which introduces some notation that we need for Theorem 2. As in Bell (2004), we first consider the equation

$$\tilde{\Delta}_S \begin{bmatrix} S_p \\ S_f \end{bmatrix} = \begin{bmatrix} S_p \\ U_f \end{bmatrix},$$

where $S_p = (S_{n+1-d_S}, \ldots, S_n)'$ and $U_f = (U_{n+1}, \ldots, U_{n+h})'$ (here the p subscript refers to the "present"). The matrix $\tilde{\Delta}_S$ is a lower triangular matrix defined by

$$\tilde{\Delta}_S = \begin{bmatrix} 1_{d_S} & 0_{d_S \times h} \\ \bar{\Delta}_S & \end{bmatrix},$$

where $\bar{\Delta}_S$ is $h \times (d_S + h)$-dimensional. Thus $\tilde{\Delta}_S$ is square of dimension $d_S + h$ and is invertible (because it is unit lower triangular). Next we compute the optimal estimate of U_f given U, which will be denoted by $\hat{U}_f$. Let the covariance matrix of (U', U_f') be subdivided into four portions:

$$\Sigma_{(U,U_f)} = \begin{bmatrix} \Sigma_U & \Sigma_{UU_f} \\ \Sigma_{U_f U} & \Sigma_{U_f} \end{bmatrix},$$

where Σ_{UU_f} denotes $\mathbb{E}[UU_f']$, and so forth. Then $\hat{U}_f$ is given by (cf. Bell, 2004, formula (12.28))

$$\hat{U}_f = \Sigma_{U_f U} \Sigma_U^{-1} U.$$

So we define the matrix D by the formula

$$D = [0_{h \times d_S} 1_h] \tilde{\Delta}_S^{-1} \begin{bmatrix} 0_{d_S \times (n-d_S)} & 1_{d_S} \\ \Sigma_{U_f U} \Sigma_U^{-1} \Delta_S & \end{bmatrix}. \tag{11}$$

This is a simple matrix approach to forecasting nonstationary processes. With this notation, we can state the main theorem for estimating future values of an unobserved component.

THEOREM 2. *Suppose that Assumption A holds for the model decomposition (1), that δ^N and δ^S share no common zeros, and that $\{U_t\}$ and $\{V_t\}$ are mean zero weakly stationary, uncorrelated with one another, and purely nondeterministic. Then the minimum MSE linear estimate of S_f given Y is $\hat{S}_f = DFY$, where F is given by (8) and D is given by (11). In addition,*

$$\begin{bmatrix} \hat{S} \\ \hat{S}_f \end{bmatrix} = \begin{bmatrix} 1_n \\ D \end{bmatrix} FY$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

has error covariance matrix

$$\begin{bmatrix} 1_n \\ D \end{bmatrix} M^{-1} [1_n D'] + \begin{bmatrix} 0 & 0 \\ 0 & G \end{bmatrix},$$

where G is the covariance matrix of $DS - S_f$, *which is given in (A.3) in the Appendix.*

Remark 3. Similar approaches can yield backcast formulas, in addition to estimates of unobserved components at time points where some of the observed data are missing. For the backcasting case, it is convenient to place the initial values at the end of the series and reverse the discussion. For both of these problems, the derivation is quite a bit more complicated.

The method of Theorem 2 allows for the forecasting of future values of an estimated signal. For example, if one is interested in future values of an unobserved trend or cycle or nonseasonal component, its optimal forecasts can be estimated in this way. A quite particular application comes from the literature on growth rates, where, for example, the growth rate over an interval p of a signal of interest is defined by $S_{t+p} - S_{t-p}$. This can be considered a "centered" measure of the rate of growth of the signal at time t (whereas $S_t - S_{t-p}$ would be an off-center measure of the growth rate of the signal at time t). In the econometrics literature, interest often focuses on growth rates at the concurrent time point n, in which case the quantity of interest is $S_{n+p} - S_{n-p}$, where data up to time n are available. This would be estimated (setting $h = p$) via

$$\hat{S}_{n+p} - \hat{S}_{n-p} = (e_{n+p} - e_{n-p})' \begin{bmatrix} 1_n \\ D \end{bmatrix} FY,$$

where e_j is the vector with jth coordinate 1 and 0 elsewhere. The error covariance matrix is easily obtained from this formula.

4. THEORETICAL APPLICATIONS OF (8)

The matrix formula (8) shows us, writing $F = Q\Delta_N$ for $Q = M^{-1}\Delta_N' \Sigma_V^{-1}$, that filtering with F first involves applying the noise differencing matrix Δ_N to the data Y. This reduces the noise component N to stationarity, but the signal will not be transformed to stationarity, because no zeros of δ^S are shared by δ^N. It is intuitive that the signal extraction filters for every time point t involve differencing the noise component and passing the signal component; of course, the signal is altered by the filtering, but its nonstationarity is preserved. Subsequently, some concrete applications of the matrix approach are given.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

4.1. Symmetry Properties

When $\{U_t\}$ and $\{V_t\}$ are weakly stationary, both F and M^{-1} have an interesting symmetry property; namely, the ith row is the reverse of row $n - i + 1$. Define the transverse transpose of a square matrix to be the matrix obtained by flipping the entries about the diagonal running from lower left to upper right. For any n-dimensional square matrix A, this transverse transpose is given by $A^*_{ij} = A_{n-j+1,n-i+1}$. It is easy to show that (i) $(AB)^* = B^*A^*$ and (ii) $(A^{-1})^* = (A^*)^{-1}$. Then the preceding symmetry property is simply the statement that $F^* = F'$ (and similarly for M^{-1}). But this is easily verified using the preceding two properties of transverse transpose, along with

$$(\Delta_N' \Sigma_V^{-1} \Delta_N)^* = \Delta_N' \Sigma_V^{-1} \Delta_N \tag{12}$$

(and similarly for $\Delta_S' \Sigma_U^{-1} \Delta_S$). This identity (12) relies on the special properties of the matrices Δ_N. Consider the $n \times n$ matrix $\tilde{\Delta}_N$ with entry ij given by δ^N_{i-j}, so that $[0_{(n-d_N)\times d_N} 1_{n-d_N}]\tilde{\Delta}_N = \Delta_N$. The dimension subscripts will be omitted for brevity. Now $\tilde{\Delta}_N$ is Toeplitz and note that $A^* = A$ for any Toeplitz matrix (as mentioned in Remark 2, (8) may hold when Σ_U and Σ_V are not Toeplitz, and in this case the preceding symmetry properties of F and M^{-1} need not be true). We can write

$$\Delta_N' \Sigma_V^{-1} \Delta_N = \tilde{\Delta}_N' \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_V^{-1} \end{bmatrix} \tilde{\Delta}_N.$$

Applying the $*$ operator yields

$$(\Delta_N' \Sigma_V^{-1} \Delta_N)^* = \tilde{\Delta}_N^* \begin{bmatrix} \Sigma_V^{-1*} & 0 \\ 0 & 0 \end{bmatrix} (\tilde{\Delta}_N')^* = \Delta_N' \Sigma_V^{-1} \Delta_N,$$

where in the last equality we have used the properties (i) and (ii) given earlier and the easily verified property that $[10]\tilde{\Delta}_N' = \Delta_N$. This verifies (12).

It follows that if n is odd, the "central" filter for row $(n + 1)/2$ is a reverse of itself, which implies that it is symmetric. Also, because M^{-1} is the covariance matrix of the signal extraction errors, it follows from (9) that the MSEs (which are the diagonal entries of M^{-1}) have the symmetry property $\mathbb{E}(\hat{S}_t - S_t)^2 = \mathbb{E}(\hat{S}_{n-t+1} - S_{n-t+1})^2$ for each t (see Figures 2 and 6). It is not clear that these symmetry properties can be easily obtained from state space algorithms.

4.2. Fundamental Zeros of the Filter Transfer Functions

The formula (8) for the signal extraction filter matrix F allows us to derive some important properties in the frequency domain. The lth row of F contains the coefficients of the filter that produces $\hat{S}_l$. The frequency response function of this filter is

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

$$H_l(\lambda) = \sum_{j=1}^{n} F_{lj} e^{-i(l-j)\lambda} = e^{-il\lambda} \sum_{j=1}^{n} F_{lj} e^{ij\lambda} \tag{13}$$

for $\lambda \in [-\pi, \pi]$ and $i = \sqrt{-1}$ (see Pollock, 1999). From H_l, one can easily obtain the gain $G_l(\lambda) = |H_l(\lambda)|$ and phase function $\phi_l(\lambda) = \tan^{-1}(\operatorname{Im} H_l(\lambda)/\operatorname{Re} H_l(\lambda))$, which is well defined when $H_l(\lambda) \neq 0$ (for a similar treatment, see Findley and Martin, 2006). Writing $F = Q\Delta_N$ as before, we have

$$H_l(\lambda) = \sum_{k=1}^{n} Q_{lk} \sum_{j=1}^{n} \delta^N_{k-j+d_N} e^{i(j-l)\lambda} = e^{-il\lambda} \sum_{k=1}^{n} Q_{lk} e^{ik\lambda} \delta^N(e^{-i\lambda}).$$

In other words, the frequency response for row l of F is given by the product of the frequency responses for $\delta^N(z)$ and row l of Q. Hence G_l is given by the product of the gain functions for $\delta^N(z)$ and row l of Q, and ϕ_l is given by the sum of the corresponding phase functions. In this manner, Appendix D of Findley and Martin (2006) computes continuous phase and phase delay functions for model-based concurrent seasonal adjustment and trend filters. In the examples that Findley and Martin (2006) consider, the phase function for $\delta^N(e^{-i\lambda})$ is simple to write down and can be continuously defined at the noise frequencies corresponding to the zeros of $\delta^N(z)$ on the unit circle. The phase function for row l of Q is well defined everywhere for many examples and hence can be separately calculated and added to the phase function of $\delta^N(e^{-i\lambda})$, yielding a continuous phase function at all frequencies.

The calculations required to produce the phase functions of Findley and Martin (2006) depend on knowing Q in the product $F = Q\Delta_N$. But Q cannot be obtained from the other approaches discussed in this paper. That is, neither the state space approach nor the method of SEATS (i.e., application of the method of Tunnicliffe–Wilson to obtain the same filter as that of the bi-infinite-sample Wiener–Kolmogorov signal extraction filter applied to data extended by forecasts and backcasts) can produce the matrix Q.

5. EMPIRICAL ILLUSTRATIONS

5.1. A Note on Efficient Computation of *F* and *M*

An alternative "square root information" approach to the formulas for F and M^{-1} is given here, which has certain computational advantages. Let $T_U = \Sigma_U^{-1/2}$ and $T_V = \Sigma_V^{-1/2}$ be square matrices obtained from Cholesky decompositions of Σ_U^{-1} and Σ_V^{-1}, respectively. Then define the $2n - (d_U + d_V)$ by $2n$-dimensional matrix T by

$$T = \begin{bmatrix} T_U \Delta_S & 0 \\ -T_V \Delta_N & T_V \Delta_N \end{bmatrix},$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

for which $T[S'Y']'$ consists of linear combinations of U and V. Next, form the QR decomposition of T (Golub and Van Loan, 1996), where Q is square and orthonormal and R is upper triangular. Writing R in block form, we have

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix}.$$

Then it is easy to show that $F = -R_{11}^{-1}R_{12}$ and $M = R_{11}'R_{11}$. Because R_{11} is upper triangular, both F and M^{-1} can be speedily computed once R is determined. The matrices T_U and T_V are readily determined (e.g., use of the Durbin–Levinson algorithm—Brockwell and Davis, 1991). The QR decomposition takes about n^3 operations and is quite stable numerically. Thus for computational purposes, this approach is to be recommended (an alternative proof of Theorem 1 can also be derived). In summary, the procedure is as follows.

1. Obtain T_U and T_V.
2. Construct T.
3. Get R from the QR decomposition of T.
4. Set $F = -R_{11}^{-1}R_{12}$ and $M = R_{11}'R_{11}$.

This approach can also be generalized to handle forecasting and backcasting.

5.2. Estimating Component Models

To implement the results of Theorem 1, it is necessary to obtain the component models for S_t and N_t from the data. Together with (2), specifying autoregressive moving average (ARMA) models for U_t and V_t will fully specify an ARIMA model for S_t and N_t. In the structural approach of Harvey (1989), the ARMA model for W_t is determined by the component models. The Gaussian likelihood for W_t depends directly on Σ_W, which in turn depends on the ARMA models for U_t and V_t. Therefore maximization of the likelihood provides estimates of the ARMA parameters for both U_t and V_t. This approach differs from the canonical decomposition technique of Hillmer and Tiao (1982), where an ARMA model for W_t is specified first; one estimates the ARMA parameters for W_t, and then one tries to decompose the model for W_t into component models, using (5).

Of course, the specification of component models is simply a basic requirement of model-based signal extraction. Oftentimes, however, the observed data are modeled as the sum of several (i.e., more than two) unobserved ARIMA components, and the signal and noise components actually consist of a partition of these various unobserved components; recall the examples (3) and (4) of Section 2. At first, it seems that one would have to determine ARIMA models for S_t and N_t from all the various ARIMA models for the unobserved components—a procedure that requires spectral factorization (Hillmer and Tiao, 1982; Burman, 1980) and a substantial amount of programming. However, the actual component models are not needed; rather only their autocovariance generating func-

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

tions are required. In SEATS one can obtain the Wiener–Kolmogorov signal extraction filter directly from a knowledge of the component models, and one need not determine an exact ARIMA model for signal and noise (though SEATS does this for informational purposes). Similarly in the matrix approach, we see that (8) only depends on the autocovariances of signal and noise, and these can be quickly obtained directly from the ARIMA models for the unobserved components, thus circumventing the need to obtain ARIMA models for S_t and N_t. We illustrate this through an example.

Consider the unobserved components model (4) of Section 2 and suppose that our signal of interest is the cycle. For concreteness, suppose that the component models are given by

$$\phi(B)\mathcal{C}_t = \theta_{\mathcal{C}}(B)\epsilon_t^{\mathcal{C}},$$

$$(1 + B + \cdots + B^{11})\mathcal{S}_t = \theta_{\mathcal{S}}(B)\epsilon_t^{\mathcal{S}},$$

$$(1 - B)^2\mathcal{T}_t = \theta_{\mathcal{T}}(B)\epsilon_t^{\mathcal{T}},$$

where the irregular $\mathcal{I}_t$ is white noise and all the ϵ_t sequences are white noise sequences (that are independent of each other). Because $\mathcal{C}_t$ is the signal, the noise is

$$N_t = \mathcal{S}_t + \mathcal{T}_t + \mathcal{I}_t.$$

The ARIMA model for the noise is determined by multiplying all differencing polynomials from its constituent parts and applying this product to N_t. Then its moving average polynomial is determined through algebra and spectral factorization techniques:

$$(1 - B)(1 - B^{12})N_t$$

$$= (1 - B)^2\theta_{\mathcal{S}}(B)\epsilon_t^{\mathcal{S}} + (1 + B + \cdots + B^{11})\theta_{\mathcal{T}}(B)\epsilon_t^{\mathcal{T}} + (1 - B)(1 - B^{12})\mathcal{I}_t. \quad \textbf{(14)}$$

The right-hand side can be reexpressed, using spectral factorization techniques, as a single moving average process, written as $\theta_N(B)\epsilon_t^N$. Now to avoid the labor of spectral factorization, observe that we only need to know the autocovariances of $V_t = (1 - B)(1 - B^{12})N_t$, not the entire ARIMA model for N_t. Reexpressing (14) in matrix form yields the equations

$$V = \Delta_N N = \underline{\Delta}_{\mathcal{T}} G + \underline{\Delta}_{\mathcal{S}} H + \Delta_N \mathcal{I},$$

$$G = \Delta_{\mathcal{S}}\mathcal{S},$$

$$H = \Delta_{\mathcal{T}}\mathcal{T}.$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

Here the differencing matrices are defined in a fashion similar to those in Section 2. Now because the vectors G, H, and $\mathcal{I}$ are uncorrelated with each other by assumption, we obtain

$$\Sigma_V = \underline{\Delta}_{\mathcal{T}} \Sigma_G \underline{\Delta}'_{\mathcal{T}} + \underline{\Delta}_{\mathcal{S}} \Sigma_H \underline{\Delta}'_{\mathcal{S}} + \Delta \Sigma_{\mathcal{I}} \Delta'.$$

Of course because $G_t = \theta_{\mathcal{S}}(B)\epsilon_t^{\mathcal{S}}$ and $H_t = \theta_{\mathcal{T}}(B)\epsilon_t^{\mathcal{T}}$, their autocovariances are easily obtained by standard algorithms (see Tunnicliffe-Wilson, 1979). In this way Σ_V is determined. This provides a flexible procedure, because if we now are interested in another signal, we can construct a new F from only a knowledge of the autocovariances of the differenced unobserved components.

More generally, suppose that the noise (though the same argument applies to the signal) can be expressed as the sum of m unobserved components $\mathcal{C}_t^{(i)}$ for $1 \le i \le m$. Suppose that each of these components satisfies $\delta^{(i)}(B)\mathcal{C}_t^{(i)} = \nu_t^{(i)}$, where $\delta^{(i)}(z)$ is a differencing polynomial and the $\nu_t^{(i)}$'s are mean zero weakly stationary processes that are uncorrelated with one another. If we know the autocovariance matrix $\Sigma_{\nu^{(i)}}$ of $\nu^{(i)}$ and there is no z such that $\delta^{(i)}(z) = 0$ for all i, then we can construct the covariance matrix for the differenced noise as follows. Let $\bar{\delta}^{(i)}(z) = \Pi_{j \ne i} \delta^{(j)}(z)$; then we have

$$V_t = \Pi_{j=1}^m \delta^{(j)}(B) N_t = \sum_{i=1}^m \bar{\delta}^{(i)}(B) \nu_t^{(i)}. \tag{15}$$

Then, letting $\bar{\Delta}^{(i)}$ be defined as previous differencing matrices but with row entries given by $\bar{\delta}^{(i)}$, we have

$$\Sigma_V = \sum_{i=1}^m \bar{\Delta}^{(i)} \Sigma_{\nu^{(i)}} \bar{\Delta}^{(i)\prime}. \tag{16}$$

The condition that there is no common zero among the $\delta^{(i)}(z)$ polynomials ensures that Σ_V is invertible. Imposing this condition is not restrictive, because the presence of such common factors could be canceled out of equation (15). If any of the m components are stationary, then the corresponding $\delta^{(i)}(z) = 1$.

5.3. Examples

We now describe some examples of the application of formula (8). Consider the series of U.S. Retail Sales of Shoe Stores data from the monthly Retail Trade Survey, from 1984 to 1998 (so $n = 170$). The logarithms of the series can be modeled with an airline model (Brockwell and Davis, 1991). Then applying (3) to the logged time series Y_t, we can obtain estimated component models for seasonal $\mathcal{S}_t$, trend $\mathcal{T}_t$, and irregular $\mathcal{I}_t$ given as follows:

$$(1 - B)(1 - B^{12}) Y_t = (1 - 0.57B)(1 - 0.34B^{12}) \epsilon_t^Y, \qquad \sigma^2_{\epsilon^Y} = 0.00096,$$

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

$$r(B)\mathcal{S}_t = (1 + 1.11B + 0.96B^2 + 0.74B^3 + 0.47B^4 + 0.20B^5 - 0.03B^6 - 0.23B^7 - 0.36B^8 - 0.47B^9 - 0.51B^{10} - 0.68B^{11})\epsilon_t^{\mathcal{S}}, \qquad \sigma^2_{\epsilon^{\mathcal{S}}} = 0.000093,$$

$$(1 - B)^2\mathcal{T}_t = (1 + 0.09B - 0.91B^2)\epsilon_t^{\mathcal{T}}, \qquad \sigma^2_{\epsilon^{\mathcal{T}}} = 0.000018,$$

$$\sigma^2_{\mathcal{I}} = 0.00026. \tag{17}$$

Here the seasonal differencing operator is $r(B) = 1 + B + \cdots + B^{11}$, and the trend differencing operator is $(1 - B)^2$. The ϵ_t sequences are white noise innovation series. The parameters for the data model were found via maximum likelihood, and the component models were found via the canonical decomposition (Hillmer and Tiao, 1982). If we wish to seasonally adjust the data, the noise is the seasonal $\mathcal{S}_t$, and the signal is the trend-irregular $\mathcal{T}_t + \mathcal{I}_t$; hence the signal differencing operator is $\delta^S(B) = (1 - B)^2$, and the noise differencing operator is $r(B)$. Then the corresponding differencing matrices are denoted by $\Delta_{\mathcal{T}}$ and $\Delta_{\mathcal{S}}$, respectively. The covariance matrices for differenced signal and noise are $\Sigma_{\Delta_{\mathcal{T}}(\mathcal{T}+\mathcal{I})}$ and $\Sigma_{\Delta_{\mathcal{S}}\mathcal{S}}$, respectively, which can be computed from (17) directly. In particular, letting $G = \Delta_{\mathcal{T}}\mathcal{T}$, we have from (16) $\Sigma_{\Delta_{\mathcal{T}}(\mathcal{T}+\mathcal{I})} = \Sigma_G + \Delta_{\mathcal{T}}\Sigma_{\mathcal{I}}\Delta'_{\mathcal{T}}$. The signal extraction matrix is given by

$$F = (\Delta'_{\mathcal{T}}\Sigma^{-1}_{\Delta_{\mathcal{T}}(\mathcal{T}+\mathcal{I})}\Delta_{\mathcal{T}} + \Delta'_{\mathcal{S}}\Sigma^{-1}_{\Delta_{\mathcal{S}}\mathcal{S}}\Delta_{\mathcal{S}})^{-1}\Delta'_{\mathcal{S}}\Sigma^{-1}_{\Delta_{\mathcal{S}}\mathcal{S}}\Delta_{\mathcal{S}},$$

where all autocovariances can be obtained from (17) via algorithms in Tunnicliffe-Wilson (1979). In Figure 1 we plot the logged series together with the estimated seasonally adjusted data. The seasonal adjustment is enclosed by error bands, given by plus and minus twice the square root of the mean squared error at each time point. Figure 2 displays the MSE curve by itself; note the increased error at the beginning and end of the series. Next, Figure 3 displays seasonal adjustment filters for three time points: $t = 86$ (denoted by "Central"), $t = 158$ (denoted by "Preliminary"), and $t = 170$ (denoted by "Concurrent"). Now by applying the material from Section 4.2, the squared gain of the filter at row l is given by

$$G_l^2(\lambda) = \left(\sum_{j=1}^{n} F_{lj}\cos(j\lambda)\right)^2 + \left(\sum_{j=1}^{n} F_{lj}\sin(j\lambda)\right)^2.$$

Plots of the squared gains over $[0, \pi]$ for the central, preliminary, and concurrent filters are given in Figure 4. For a discussion of the gain and phase properties of finite-sample signal extraction filters, see Findley and Martin (2006). As Section 4.2 establishes, the squared gains are zero at the zeros of $r(e^{-i\lambda})$, which for $\lambda \geq 0$ are at $2\pi k/12$ for $k = 1, 2, \ldots, 6$. Thus the filters suppress variance components with periods of 12, 6, 4, 3, 2.4, and 2 months.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

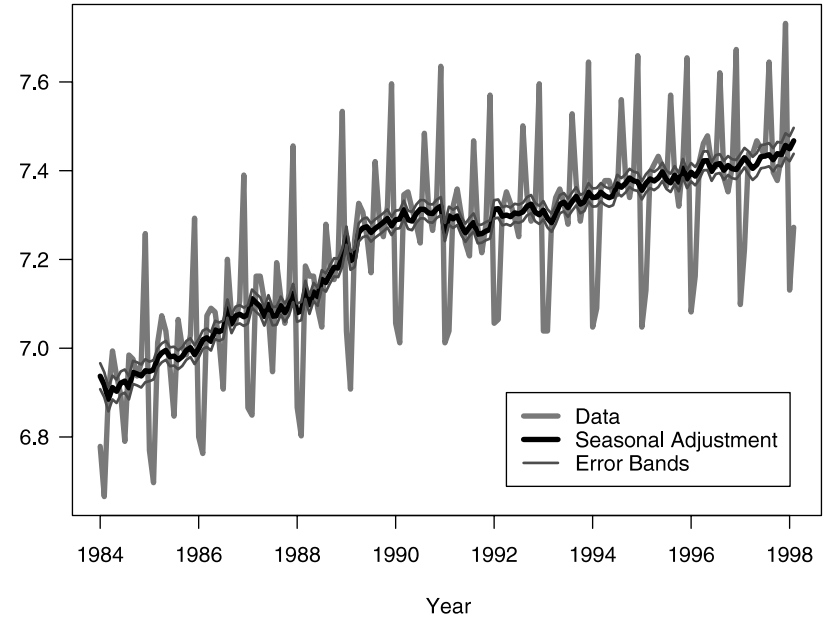


FIGURE 1. Logarithm of U.S. Retail Sales of Shoe Stores with seasonal adjustment. Error bands are given by plus and minus 2 standard deviations.

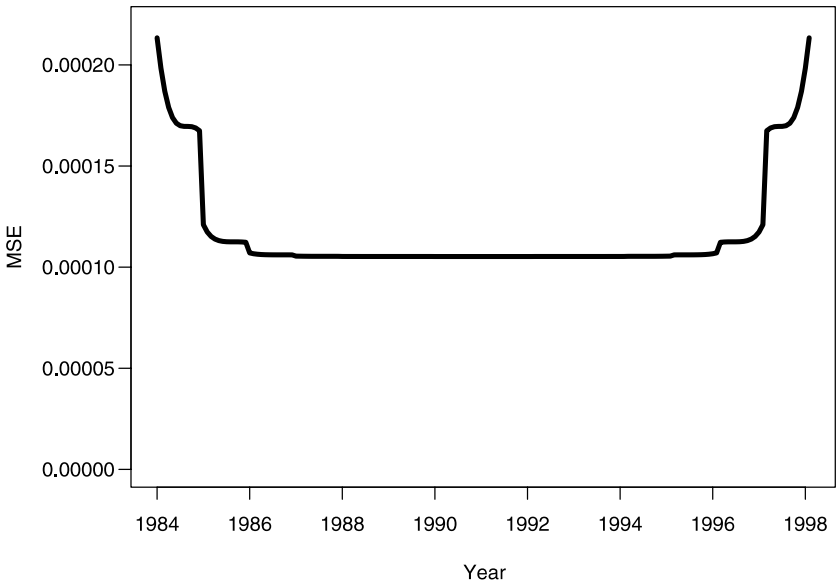


FIGURE 2. MSE of the seasonal adjustment as a function of time in the sample.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

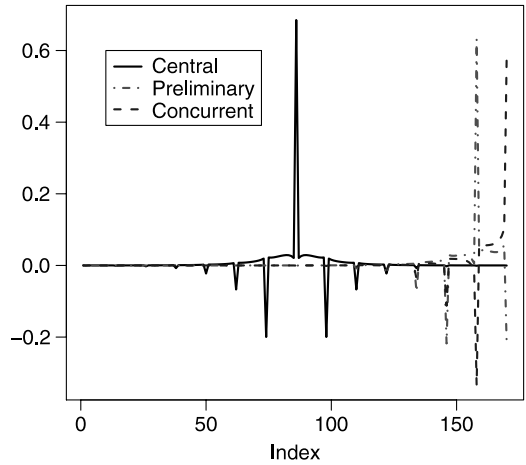


FIGURE 3. Seasonal adjustment filter coefficients for central, preliminary, and concurrent filters.

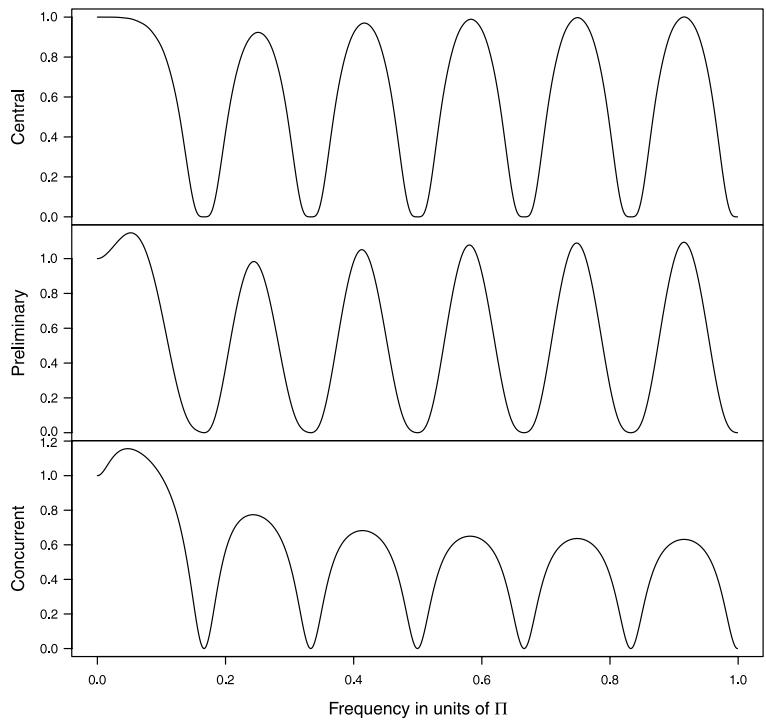


FIGURE 4. Squared gain plots for seasonal adjustment filters.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

As a second example we consider the same series, but now our signal of interest is the trend $\mathcal{T}_t$, so that the noise is the seasonal-irregular $\mathcal{S}_t + \mathcal{I}_t$. If we wish to estimate the trend the signal differencing operator is $(1 - B)^2$, and the noise differencing operator is $r(B)$. Then the corresponding differencing matrices are denoted by $\Delta_{\mathcal{T}}$ and $\Delta_{\mathcal{S}}$, respectively. The covariance matrices for differenced signal and noise are $\Sigma_{\Delta_{\mathcal{T}}\mathcal{T}}$ and $\Sigma_{\Delta_{\mathcal{S}}(\mathcal{S}+\mathcal{I})}$, respectively, which can be computed from (17). Indeed, letting $H = \Delta_{\mathcal{S}}\mathcal{S}$ and using (16) we have $\Sigma_{\Delta_{\mathcal{S}}(\mathcal{S}+\mathcal{I})} = \Sigma_H + \Delta_{\mathcal{S}}\Sigma_{\mathcal{I}}\Delta'_{\mathcal{S}}$. The signal extraction matrix is given by

$$F = (\Delta'_{\mathcal{T}}\Sigma^{-1}_{\Delta_{\mathcal{T}}\mathcal{T}}\Delta_{\mathcal{T}} + \Delta'_{\mathcal{S}}\Sigma^{-1}_{\Delta_{\mathcal{S}}(\mathcal{S}+\mathcal{I})}\Delta_{\mathcal{S}})^{-1}\Delta'_{\mathcal{S}}\Sigma^{-1}_{\Delta_{\mathcal{S}}(\mathcal{S}+\mathcal{I})}\Delta_{\mathcal{S}},$$

where all autocovariances can be obtained from (17) via algorithms in Tunnicliffe-Wilson (1979). In Figure 5 we plot the logged series together with the trend estimate. This estimate is enclosed by error bands, given by plus and minus twice the square root of the MSE at each time point. Figure 6 displays the MSE curve alone. Next, Figure 7 displays central, preliminary, and concurrent trend filters for three time points. As in the previous example, we can compute the squared gains, which are displayed in Figure 8.

In comparing Figures 1 and 5, observe that the trend estimate is much smoother than the seasonally adjusted estimate; this is accounted for by the presence of the irregular component in the seasonally adjusted component. The symmetry

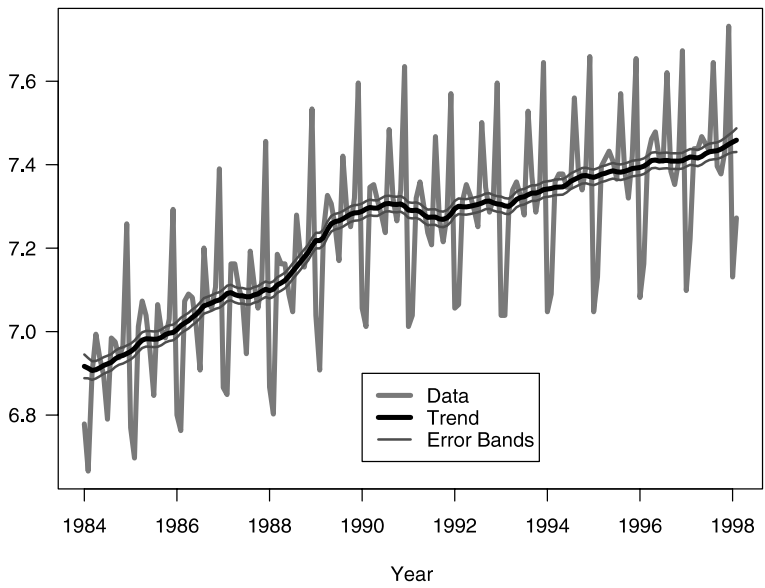


FIGURE 5. Logarithm of U.S. Retail Sales of Shoe Stores with trend estimate. Error bands are given by plus and minus 2 standard deviations.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

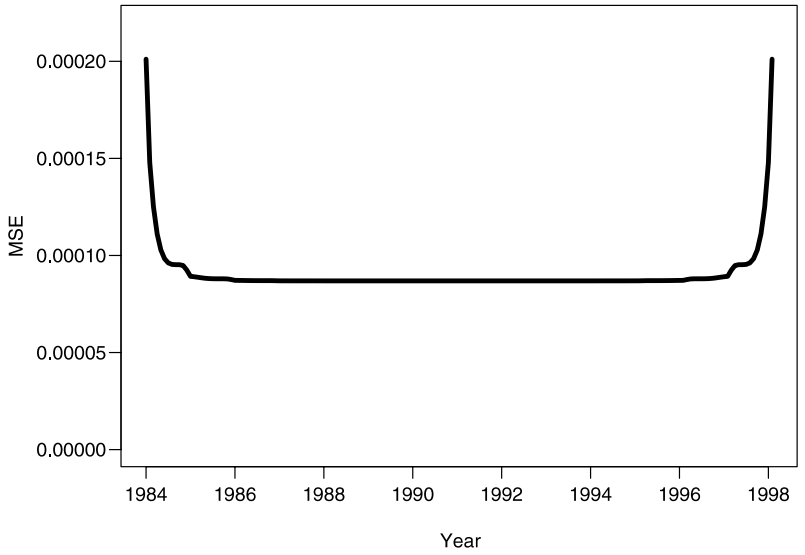


FIGURE 6. MSE of the trend estimate as a function of time in the sample.

of the mean squared errors is apparent from Figures 2 and 6. The upturning of the mean squared error at the beginning and end of the sample is a well-known effect and is essentially due to the increased error in estimating the signal at the boundary of the sample. In both Figures 3 and 7 we see that the filter coeffi-

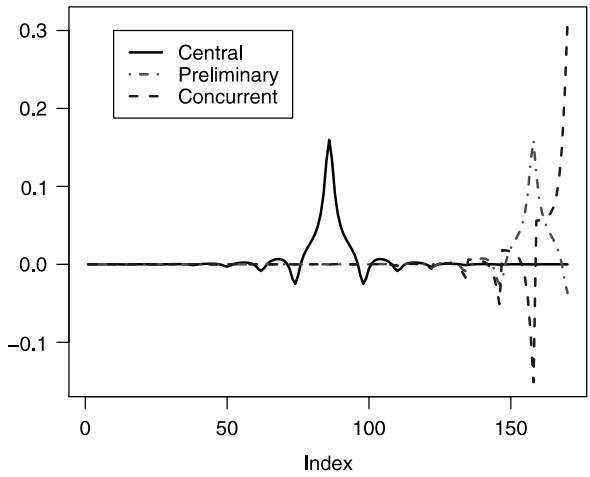


FIGURE 7. Trend filter coefficients for central, preliminary, and concurrent filters.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

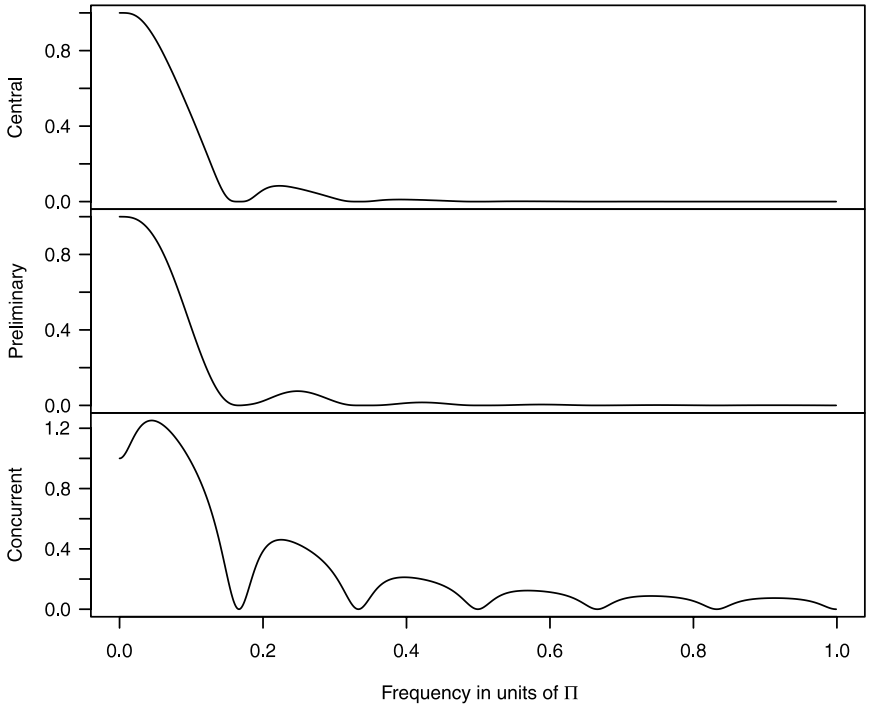


FIGURE 8. Squared gain plots for trend filters.

cients contain negative spikes at seasonal lags, which effect the suppression of the seasonal component. The greater width of the central spike in the trend filters of Figure 7 allows for inclusion of the trend component and exclusion of the irregular component. Finally, the squared gain plots of Figures 4 and 8 show the attenuation of seasonal frequencies; for the trend filters, this is combined with a low-pass filter effect. In comparing central, preliminary, and concurrent filters, there is some variation in trough width among the squared gain plots. For further study of such finite-sample filters and their frequency domain properties, see Findley and Martin (2006).

REFERENCES

Bell, W. (1984) Signal extraction for nonstationary time series. *Annals of Statistics* 12, 646–664.

Bell, W. (2004) On RegComponent time series models and their applications. In Andrew C. Harvey, Siem Jan Koopman, & Neil Shephard (eds.), *State Space and Unobserved Component Models: Theory and Applications*, pp. 248–283. Cambridge University Press.

Bell, W. & S. Hillmer (1988) A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models. SRD Research report RR-88/22, Bureau of the Census. http://www.census.gov/srd/papers/pdf/rr88-22.pdf.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

Bell, W. & S. Hillmer (1991) Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis* 12, 283–300.
Bell, W. & S. Hillmer (1992) Correction to "Initializing the Kalman filter for nonstationary time series models." *Journal of Time Series Analysis* 13, 281–282.
Bell, W. & D. Martin (2004) Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. *Journal of Time Series Analysis* 25, 603–625.
Brockwell, P. & R. Davis (1991) *Time Series: Theory and Methods.* Springer-Verlag.
Burman, J. (1980) Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A* 143, 321–337.
Cleveland, W. & G. Tiao (1976) Decomposition of seasonal time series: A model for the Census X-11 program. *Journal of the American Statistical Association* 71, 581–587.
de Jong, P. & M. MacKinnon (1988) Covariances for smoothed estimates in state space models. *Biometrika* 75, 601–602.
Durbin, J. & S. Koopman (2001) *Time Series Analysis by State Space Methods.* Oxford University Press.
Durbin, J. & B. Quenneville (1997) Benchmarking by state space methods. *International Statistical Review* 65, 23–48.
Findley, D. & D. Martin (2006) Frequency domain analyses of SEATS and X-11/12-ARIMA seasonal adjustment filters for short and moderate-length time series. *Journal of Official Statistics* 22, 1–34.
Findley, D., T. McElroy, & K. Wills (2004) Diagnostics for ARIMA-model-based seasonal adjustment. In *2004 Proceedings of the American Statistical Association, Section on Business and Economics Statistics*. American Statistical Association, 1187–1194 (CD-ROM).
Golub, G. & C. Van Loan (1996) *Matrix Computations.* Johns Hopkins University Press.
Gómez, V. & A. Maravall (2001) Seasonal adjustment and signal extraction in economic time series. In D. Peña, G. Tiao, & R. Tsay (eds.), *A Course in Time Series Analysis.* Wiley.
Hannan, E. (1967) Measurement of a wandering signal amid noise. *Journal of Applied Probability* 4, 90–102.
Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press.
Hillmer, S. & G. Tiao (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* 77, 63–70.
Kailath, T., A. Sayed, & B. Hassibi (2000) *Linear Estimation.* Prentice-Hall.
Koopman, S. & A. Harvey (2003) Computing observation weights for signal extraction and filtering. *Journal of Economic and Dynamic Control* 27, 1317–1333.
Maravall, A. & G. Caporello (2004) Program TSW: Revised Reference Manual. Working paper 2004, Research Department, Bank of Spain. http://www.bde.es.
McElroy, T. (2005) Matrix Formulas for Nonstationary Signal Extraction. SRD Research report RR-2005/04, Bureau of the Census. http://www.census.gov/srd/papers/pdf/rrs2005-04.pdf.
McElroy, T. & A. Sutcliffe (2006) An iterated parametric approach to nonstationary signal extraction. *Computational Statistics and Data Analysis, Special Issue on Signal Extraction* 50, 2206–2231.
Nguyen, T., W. Bell, & J. Gomish (2002) Investigating model-based time series methods to improve estimates from monthly value of construction put-in-place surveys. *Proceedings of the American Statistical Association, Section on Survey Research Methods*. American Statistical Association, 2470–2475 (CD-ROM).
Pollock, S. (1999) *A Handbook of Time-Series Analysis, Signal Processing, and Dynamics.* Academic Press.
Pollock, S. (2001) Filters for short non-stationary sequences. *Journal of Forecasting* 20, 341–355.
Sobel, E. (1967) Prediction of a noise-distorted, multivariate, non-stationary signal. *Journal of Applied Probability* 4, 330–342.
Tunnicliffe-Wilson, G. (1979) Some efficient computational procedures for high order *ARMA* models. *Journal of Statistical Computation and Simulation* 8, 301–309.

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

APPENDIX

Proof of Theorem 1. First note that all covariance matrices for U, V, and W are invertible because the stationary series $\{U_t\}$, $\{V_t\}$, and $\{W_t\}$ are assumed to be purely nondeterministic—see Proposition 5.1.1 of Brockwell and Davis (1991). Let 1_n denote the n-dimensional identity matrix and let F have the stated form (8). First we show that M is invertible by showing that $Mx = 0$ implies that $x = 0$. If a vector x were in the null space of M, then because of its symmetry and nonnegative definiteness, x would have to be in the null space of both Δ_S and Δ_N. As demonstrated in Lemma 2 of McElroy and Sutcliffe (2006), x must be zero because $\delta^S(z)$ and $\delta^N(z)$ have no common zero. Originally, we derived (8) from formulas of Bell and Hillmer (1988). Here, to save space we simply verify that $\hat{S}$ defined by (8) has the defining property of the optimal estimate, namely, that the error process $\epsilon = \hat{S} - S$ is orthogonal to all linear combinations of the observations Y (the original proof is in McElroy, 2005). Now

$$\epsilon = FY - S = (F - 1_n)S + FN = M^{-1}\Delta_N'\Sigma_V^{-1}V - M^{-1}\Delta_S'\Sigma_U^{-1}U \tag{A.1}$$

and therefore is uncorrelated with the initial values Y_* by Assumption A. In addition, by (7) and (6) we have

$$\mathbb{E}[\epsilon W'] = M^{-1}\Delta_N'\Sigma_V^{-1}\Sigma_V\underline{\Delta}_S' - M^{-1}\Delta_S'\Sigma_U^{-1}\Sigma_U\underline{\Delta}_N' = M^{-1}(\Delta' - \Delta') = 0.$$

Because Y can be expressed in terms of its initial values Y_* and W (see e.g., Bell, 2004), this shows that ϵ is orthogonal to any linear combination of Y. Moreover, we can use (A.1) to compute $\mathbb{E}[\epsilon\epsilon'] = M^{-1}$. Hence the theorem is proved. ■

Proof of Theorem 2. We will work with the full estimated component $(\hat{S}', \hat{S}_f')$. As in the proof of Theorem 1, it suffices to demonstrate that the error process is orthogonal to Y. Now from (A.1) and (10)

$$\epsilon = \begin{bmatrix} \hat{S} \\ \hat{S}_f \end{bmatrix} - \begin{bmatrix} S \\ S_f \end{bmatrix},$$

$$= \begin{bmatrix} 1_n \\ D \end{bmatrix} M^{-1}\Delta_N'\Sigma_V^{-1}V - \begin{bmatrix} 1_n \\ D \end{bmatrix} M^{-1}\Delta_S'\Sigma_U^{-1}U + \begin{bmatrix} 0 \\ DS - S_f \end{bmatrix}. \tag{A.2}$$

We next compute the error $DS - S_f$, using the partition

$$\tilde{\Delta}_S^{-1} = \begin{bmatrix} 1_{d_S} & 0_{d_S \times h} \\ A & B \end{bmatrix}.$$

Here B is a square matrix of dimension h and lower triangular, with entries given by the coefficients of $1/\delta_S(z)$ (see Bell, 2004). Hence we have

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press

$$
\begin{aligned}
DS - S_f &= [0_{h\times d_S} 1_h]\left(\tilde{\Delta}_S^{-1}\begin{bmatrix} S_p \\ \Sigma_{U_f U}\Sigma_U^{-1}\Delta_S S\end{bmatrix} - \begin{bmatrix} S_p \\ S_f\end{bmatrix}\right)\\
&= [0_{h\times d_S} 1_h]\tilde{\Delta}_S^{-1}\left(\begin{bmatrix} S_p \\ \hat{U}_f\end{bmatrix} - \tilde{\Delta}_S\begin{bmatrix} S_p \\ S_f\end{bmatrix}\right)\\
&= [AB]\left(\begin{bmatrix} S_p \\ \hat{U}_f\end{bmatrix} - \begin{bmatrix} S_p \\ \bar{\Delta}_S S_f\end{bmatrix}\right)\\
&= [AB]\begin{bmatrix} 0 \\ \hat{U}_f - U_f\end{bmatrix}\\
&= B(\hat{U}_f - U_f).
\end{aligned}
$$

Note that if the initial values of S, say, S_*, were assumed to be uncorrelated with U, then the preceding calculation shows that DS is the optimal linear estimate of S_f. However, this assumption on S_* is typically not true in our context, and so we cannot conclude that DS is an optimal estimate of S_f given S. However, by Assumption A it is uncorrelated with both Y_* and W, and thus DS is an optimal estimate of S_f given Y. Now (A.2) shows that each of the three terms is mutually orthogonal: as in the proof of Theorem 1, W is orthogonal to the sum of the first two terms, and in addition W is orthogonal to $DS - S_f$ because $\hat{U}_f - U_f$ only depends on $\{U_t\}$ and is orthogonal to U. Thus W is uncorrelated with the error process. Finally, because ϵ is a linear function of only U_t and V_t, it is also uncorrelated with Y_* by Assumption A. Because the error process is orthogonal to Y, the filter must be optimal.

For the MSE calculation, we compute $\mathbb{E}[\epsilon\epsilon']$. The only difficulty is computing the cross-covariance between the second and third terms of (A.3). But because $DS - S_f = B(\hat{U}_f - U_f)$, we see that U is uncorrelated with $DS - S_f$, so that the second and third terms of (A.3) are orthogonal. Hence the error covariance matrix has the stated form, with

$$G = B(\Sigma_{U_f} - \Sigma_{U_f U}\Sigma_U^{-1}\Sigma_{UU_f})B', \tag{A.3}$$

utilizing (12.30) of Bell (2004). This concludes the proof. ■

https://doi.org/10.1017/S0266466608080389 Published online by Cambridge University Press