Research Article

Tucker S. McElroy* and Agustin Maravall

Optimal Signal Extraction with Correlated Components

Abstract: While it is typical in the econometric signal extraction literature to assume that the unobserved signal and noise components are uncorrelated, there is nevertheless an interest among econometricians in the hypothesis of hysteresis, i.e. that major movements in the economy are fundamentally linked. While specific models involving correlated signal and noise innovation sequences have been developed and applied using state space methods, there is no systematic treatment of optimal signal extraction with correlated components. This paper provides the mean square error optimal formulas for both finite samples and bi-infinite samples and furthermore relates these filters to the more well-known Wiener–Kolmogorov (WK) and Beveridge–Nelson (BN) signal extraction formulas in the case of ARIMA component models. Then we obtain the result that the optimal filter for correlated components can be viewed as a weighted linear combination of the WK and BN filters. The gain and phase functions of the resulting filters are plotted for some standard cases. Some discussion of estimation of hysteretic models is presented, along with empirical results on an economic time series. Comparisons are made between signal extractions from traditional WK filters and those arising from the hysteretic models.

Keywords: ARIMA, nonstationary, seasonality, time series

***Corresponding author: Tucker S. McElroy,** U.S. Census Bureau – CSRM, 4600 Silver Hill Road, Washington, DC 20233, USA, E-mail: tucker.s.mcelroy@census.gov
Agustin Maravall, Bank of Spain, Madrid, Spain, E-mail: maravall@bde.es

1 Introduction

In the econometric literature on nonstationary signal extraction, two popular approaches have developed, each based on an underlying assumption on the relationship of signal and noise: namely, that these components are uncorrelated (i.e. orthogonal) or that they are fully correlated (i.e., collinear). These assumptions are motivated alternatively by economic considerations – reflecting *a priori* beliefs on the nature of the driving forces underpinning an economic variable – or through statistical considerations of identifiability of a given time series model. The ambiguity is due to the fact that neither signal nor noise is actually observed, so the actual correlation structure of the putative components cannot be directly measured. In between these two extreme viewpoints – orthogonality on the one hand and collinearity on the other hand – exists a generic formulation where there is some degree of cross-correlation between signal and noise. Ghysels (1987) investigated this phenomenon for time series with trend and seasonality. The term *hysteresis* has been introduced (Jäger and Parkinson 1994) to describe this situation with respect to trends and cycles; we use it more generally in this paper to denote a correlation between the unobserved signal and noise.

More generally, the term "hysteresis" refers to a dynamical system whose output depends not only upon the inputs but upon the internal state of the system. There is a considerable amount of economic literature on this topic. Proietti (2006) gives an overview with a focus on the case of nonzero correlation; the key reference for the full correlation case is Beveridge and Nelson (1981), but also see Snyder (1985), Ghysels (1987), Ord, Koehler, and Snyder (1997), Hyndman et al. (2002), and Oh, Zivot, and Creal (2008). The literature on orthogonal components goes back to Wiener (1949), but more recent references include Bell and Hillmer (1984), Bell (1984), Harvey (1989), Harvey and Jäger (1993), Maravall (1995), and McElroy (2008). This is not an exhaustive list: for further reading see the references in Proietti (2006) and Bell (2004). The paper at hand does not seek to enter the argument as to whether hysteresis exists and/or is a useful concept, because this has been argued in many other papers. Instead we focus on providing a complete and general mathematical analysis of hysteresis, in order to better elucidate the properties of signal extraction in this context.

In particular, we provide exact optimal formulas for nonstationary signal extraction (with a nonstationary noise component) when cross-correlation is present in varying degrees, for both finite samples and bi-infinite samples. Although formulas for stationary bi-infinite samples can be found in Whittle (1963), our formulas for the nonstationary case are novel. The new finite sample formulas allow for a quite general treatment of hysteresis and are practical for implementation on real series (without requiring the labor of state space methods). Here optimality refers to a statistic that has minimum mean squared error among all statistics linear in the time series, or alternatively one that has minimum mean squared error among all statistics when the series is a sample from a Gaussian process. So the statistics are classical, in the sense that they are conditional expectations under a Gaussian assumption, but are derived under the new and somewhat heterodox assumption of cross-correlation in the components, thereby generalizing the standard formulas employed in the orthogonal paradigm.

In order for these results to be useful in applied econometric analysis, one needs algorithms to fit hysteretic models to time series data. The most viable approach at present is to set up a structural model for the observed time series, directly estimating the cross-correlation of signal and noise innovations along with the other parameters; see Proietti (2006) for an in-depth discussion. Some supplementary explication is offered in our article, with algorithms for a simple hysteretic structural model. Our main focus is to provide explicit signal extraction formulas and illustrate how these can be implemented for linear structural models.[1] Furthermore, we develop a fundamental interpretation of the optimal hysteretic filter as a convex combination of the Beveridge–Nelson (BN) and Wiener–Kolmogorov (WK) filters[2] in the case of ARIMA models, building on the classical approach of Bell and Martin (2004). The filters are then examined in the frequency domain through the plotting of their gain and phase delay functions.

The organization of this paper is as follows. In Section 2, we develop the main mathematical material, with proofs in the Appendix. The special case of ARIMA component models is developed in Section 3, and we connect the BN and WK filters to the hysteretic filters. A discussion of implementation of the signal extraction filters is also provided,[3] with a frequency domain analysis of the filters that provides further insight into the role of hysteresis. Section 4 proposes a simplistic structural hysteretic model, with a discussion of its estimation utilizing unconstrained optimization of the exact Gaussian likelihood. Then Section 5 provides an application of these models to several economic time series, with comparisons of traditional and hysteretic signal statistics. Section 6 summarizes our findings, with supplementary material in the Appendix.

---

1 By this we mean any unobserved component time series model – see Gersch and Kitagawa (1983) and Harvey (1989) – where each component, once suitably differenced to reduce to stationarity, can be viewed as a linear process.

2 Technically, the WK filter only applies to the case of stationary signal and noise, but by a standard abuse of terminology we extend this appellation to the case of nonstationary signal and noise as developed in Bell (1984).

3 R code (R Development Core Team 2012) for the finite sample filters is available from the first author.

2 Signal extraction formulas

Consider a sample of size n from a time series $\{Y_t\}$, i.e. a realization from the vector $Y = (Y_1, Y_2, \ldots, Y_n)'$. We suppose this vector can be decomposed into signal S and noise N, via $Y = S + N$. For example, the signal might be the trend component, while the noise might consist of the seasonal and irregular components. From now on, we refer to $\{Y_t\}$ as the aggregate process. Following Bell (1984), we let $\{Y_t\}$ be an integrated process such the differenced series $\{W_t\}$ is stationary, where $W_t = \delta(B)Y_t$, B is the backshift operator, and δ 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 it can be factored into relatively prime polynomials $\delta^S(z)$ and $\delta^N(z)$[i.e. polynomials with no common zeroes; see discussion in Bell (1984)], such that the series $\{U_t\}$ and $\{V_t\}$ defined via

$$U_t = \delta^S(B)S_t \qquad V_t = \delta^N(B)N_t \tag{1}$$

are mean zero stationary time series that *could possibly be correlated* 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 the orders of δ^S and δ^N; since the latter operators are relatively prime, $\delta = \delta^S \cdot \delta^N$ and $d = d_S + d_N$.

Now we can write eq. [1] in a matrix form, as follows. Let Δ be a $(n-d) \times n$ matrix with entries given by $\Delta_{ij} = \delta_{i-j+d}$ (the convention being that $\delta_k = 0$ if $k < 0$ or $k > d$), i.e.

$$\Delta = \begin{bmatrix} \delta_d & \cdots & \delta_1 & 1 & 0 & 0 & \cdots \\ 0 & \delta_d & \cdots & \delta_1 & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \delta_d & \cdots & \delta_1 & 1 \end{bmatrix}$$

The matrices Δ_S and Δ_N have entries given by the coefficients of δ^S and δ^N, 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 W, U, V, S, and N are column vectors of appropriate dimension, with covariance matrices Γ_W, Γ_U, and so forth. We also have need of differencing matrices of reduced dimension, denoted as $\underline{\Delta}_N$ and $\underline{\Delta}_S$; these have the same entries as Δ_N and Δ_S, but have dimension $(n-d) \times (n-d_S)$ and $(n-d) \times (n-d_N)$, respectively, such that

$$\Delta = \underline{\Delta}_S \Delta_N = \underline{\Delta}_N \Delta_S. \qquad [2]$$

A proof of eq. [2] can be found in McElroy and Sutcliffe (2006). With these notations it is possible to write the mean square error (MSE) linear optimal statistic $\widehat{S}$ as a linear function of Y, i.e. $\widehat{S} = FY$. The error covariance matrix, i.e. the covariance matrix of $\widehat{S} - S$, is denoted by M.

The difference from the conventional signal extraction literature is that we allow for cross-correlation between differenced signal $\{U_t\}$ and differenced noise $\{V_t\}$, namely $\mathbb{E}[UV'] = \Gamma_{UV}$ can be nonzero. In general, Γ_{UV} will not be square. Note that $\Gamma_{VU} = \mathbb{E}[VU'] = \Gamma'_{UV}$. Now from eq. [2] it follows that

$$W = \underline{\Delta}_N U + \underline{\Delta}_S V. \qquad [3]$$

Then the covariance matrices are related by

$$\Gamma_W = \underline{\Delta}_S \Gamma_V \underline{\Delta}'_S + \underline{\Delta}_N \Gamma_U \underline{\Delta}'_N + \underline{\Delta}_N \Gamma_{UV} \underline{\Delta}'_S + \underline{\Delta}_S \Gamma_{VU} \underline{\Delta}'_N, \qquad [4]$$

which follows from eq. [3]. We need to assume that Γ_U, Γ_V, and Γ_W are invertible; by stationarity of the underlying process they will be Toeplitz. As was established for conventional signal extraction theory (Bell 1984) and extended to the finite sample case by McElroy (2008), it is useful to consider an assumption relating the initial values of the process to the differenced signal and noise:

Assumption A (Bell 1984). The initial values $Y^* = (Y_1, Y_2, \ldots, Y_d)'$ are uncorrelated with $\{U_t\}$ and $\{V_t\}$.

This is Assumption A of Bell (1984), which has become a standard working assumption in state space approaches to signal extraction [see Bell and Hillmer (1991)]. A consequence of this assumption, which is appealing (see the proof of Theorem 1), is that the minimum MSE linear estimator of differenced signal U given Y is identical with the optimal MSE estimator of U given W. The next result presents a formula for the minimum MSE linear estimator of signal S given Y. In the Appendix, we present two proofs – one that verifies optimality of the estimator, and another longer proof that constructively derives the estimator.

Theorem 1 *Assume that $\{U_t\}$ and $\{V_t\}$ are mean zero stationary time series, and assume the invertibility of Γ_U, Γ_V, and Γ_W. Under Assumption A, the MSE linear optimal signal extraction filter is given by*

$$F = M^{-1}\left[\Delta'_N \Gamma_V^{-1} \Delta_N + P \Gamma_W^{-1} \Delta\right],$$

where M is an invertible matrix, and both M and P are given by

$$M = \Delta_S' \Gamma_U^{-1} \Delta_S + \Delta_N' \Gamma_V^{-1} \Delta_N$$

$$P = \Delta_S' \Gamma_U^{-1} \Gamma_{UV} \underline{\Delta}_S' - \Delta_N' \Gamma_V^{-1} \Gamma_{VU} \underline{\Delta}_N'.$$

Let the signal extraction error be denoted by $\varepsilon = FY - S$. *Then the error covariance matrix* Γ_ε *is*

$$\Gamma_\varepsilon = M^{-1} - M^{-1}\left(P\Gamma_W^{-1}P' + \Delta_N'\Gamma_V^{-1}\Gamma_{VU}\Gamma_U^{-1}\Delta_S + \Delta_S'\Gamma_U^{-1}\Gamma_{UV}\Gamma_V^{-1}\Delta_N\right)M^{-1}.$$

Remark 1 In the special case that U and V are not cross-correlated, $P = 0$ and we at once obtain the more classical signal extraction filters (McElroy 2008). When P is nonzero, F and Γ_ε need not be centro-symmetric.

Remark 2 The result also holds more generally when $\{U_t\}$ and/or $\{V_t\}$ are not stationary (e.g. they are heteroscedastic). When the latent processes are stationary, simpler formulas are available; see the proof of Theorem 1 for more detail.

For large sample sizes, the inversion of M that is required by Theorem 1 can be time-consuming and prone to rounding errors. In the Appendix, we develop an alternative formula, which is less compactly expressed than that given above, but has some computational advantages. As the sample size becomes quite large, the central filters (i.e. the middle rows of F) will have coefficients similar to those of a bi-infinite filter. The case of a bi-infinite sample is somewhat easier to describe; now the signal extraction filter is allowed to depend on past and future values of the aggregate process. In this case, we seek the MSE optimal statistic $\widehat{S}_t = \Psi(B)Y_t$, for a signal extraction filter $\Psi(z)$. Because we seek to derive time-invariant filters, we need to assume that $\{U_t\}$ and $\{V_t\}$ are jointly weakly stationary (so Remark 2 above no longer applies). We employ the notation f_W and f_U for the spectral density functions of $\{W_t\}$ and $\{U_t\}$, while f_{UV} is the cross-spectral density of $\{U_t\}$ and $\{V_t\}$– so $f_{UV}(\lambda) = f_{VU}(-\lambda)$. We adopt the conventions of Brockwell and Davis (1991), so that $\mathbb{E}[U_{t+h}V_t]$ is the lag h value of cross-covariance function (of U with V), with corresponding cross-spectral density $f_{UV}(\lambda) = \sum_{h\in\mathbb{Z}} \mathbb{E}[U_{t+h}V_t]e^{-ih\lambda}$.

In the expressions below, we utilize the abbreviation $z = e^{-i\lambda}$, it being understood by context when z refers to a power series' argument, and when it refers to the complex exponential. Then we have the following relations in analogy with eq. [4]:

$$\begin{aligned} f_W(\lambda) = f_U(\lambda)\delta^N(z)\delta^N(\bar{z}) + f_{UV}(\lambda)\delta^N(z)\delta^S(\bar{z}) \\ + f_{VU}(\lambda)\delta^S(z)\delta^N(\bar{z}) + f_V(\lambda)\delta^S(z)\delta^S(\bar{z}) \end{aligned} \qquad [5]$$

for $\lambda \in [-\pi, \pi]$. Then the frequency response function (frf) for $\Psi(z)$ is given in the following result.

Theorem 2 *Under Assumption A, and assuming that $f_W(\lambda)$is positive for all λ, the MSE linear optimal signal extraction filter has frf given by*

$$\Psi(z) = \frac{f_U(\lambda)\delta^N(z)\delta^N(\bar{z}) + f_{UV}(\lambda)\delta^N(z)\delta^S(\bar{z})}{f_W(\lambda)}.$$

Let the signal extraction error process be denoted by $\{\varepsilon_t\}$, with $\varepsilon_t = \Psi(B)Y_t - S_t$. Then the error process is weakly stationary and has spectral density

$$f_\varepsilon(\lambda) = \frac{f_U(\lambda)f_V(\lambda) - f_{VU}(\lambda)f_{UV}(\lambda)}{f_W(\lambda)}.$$

Remark 3 We can factor the numerator of $\Psi(z)$ and write the filter as

$$\Psi(z) = \frac{f_U(\lambda)\delta^N(\bar{z}) + f_{UV}(\lambda)\delta^S(\bar{z})}{f_W(\lambda)}\ \delta^N(z).$$

This shows that the filter applies the noise-differencing filter $\delta^N(z)$ first to the input series, which is a necessary aspect of a signal extraction filter where nonstationary noise is involved [see discussion in McElroy (2012)]. In the classical case wherein $f_{UV} \equiv 0$, the numerator also contains the factor $\delta^N(\bar{z})$, so that input series are initially differenced by the filter $\delta^N(z)\delta^N(\bar{z})$. For example, if $\delta^N(z) = 1 - z$ and $f_{UV} \not\equiv 0$, then the filter differences an input time series once. But if $f_{UV} \equiv 0$, the filter involves the factor $\delta^N(z)\delta^N(\bar{z}) = -\bar{z}(1-z)^2$, so that the filter involves two temporal differences.

Remark 4 An intuitive connection between the finite sample (Theorem 1) and bi-infinite sample (Theorem 2) formulas can be provided by considering the minimum MSE estimator of U given Y – see the proof of Theorem 1:

$$\Delta_S F = \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \Gamma_W^{-1} \Delta.$$

Replacing covariance matrices by the corresponding spectral densities and cross-spectral densities and differencing matrices by differencing polynomials (transposition corresponds to conjugation of the complex exponential argument) then yields

$$\left(f_U(\lambda)\,\delta^N(\bar{z}) + f_{UV}(\lambda)\,\delta^S(\bar{z})\right)\delta(z)/f_W(\lambda),$$

which is $\delta^S(z)\Psi(z)$; this is the frf for the optimal bi-infinite filter for estimating U_t given $\{Y_t\}$. The connection between signal estimators of S given Y and S_t given $\{Y_t\}$ is less direct, but can be inferred from the above relationship for the differenced signal.

It sometimes occurs in applications that f_W is noninvertible. So long as any zeroes occurring in f_W are also present in all of the components, Theorem 2 is still valid. For example, the so-called canonical trend process $\{W_t\}$ is given by $(1-B)W_t = (1+B)a_t$, where $\{a_t\}$ is white noise, and therefore f_W is noninvertible. When no hysteresis is present, the decomposition into long-term trend and cycle, as described in Kaiser and Maravall (2005), produces component models that each contain the $1+B$ factor in their moving averages. Hence $(1+z)(1+\bar{z})$ can be canceled from numerator and denominator of the frf, so that the simplified frf is well-defined.

3 The ARIMA case: relation to BN and WK filters

In this section, we consider a bi-infinite process that can be represented in terms of ARIMA component models and compare the hysteretic filters to the BN and WK filters. With notation borrowed from Bell and Martin (2004), we suppose the process satisfies $Y_t = S_t + N_t$ as before, with

$$\varphi(B)Y_t = \theta(B)\,a_t \qquad \varphi^S(B)S_t = \theta^S(B)\,b_t \qquad \varphi^N(B)N_t = \theta^N(B)\,c_t. \qquad [6]$$

The polynomials φ, φ^S, and φ^N can always be factored into portions involving roots on the unit circle and roots outside the unit circle. The former are denoted by δ, δ^S, and δ^N as in Section 2, but the latter are denoted by ϕ, ϕ^S, and ϕ^N, i.e. these are the autoregressive polynomials. The innovation sequences $\{a_t\}$, $\{b_t\}$, and $\{c_t\}$ are white noise of variance σ_a^2, σ_b^2, and σ_c^2, but we allow for the possibility that $\{b_t\}$ and $\{c_t\}$ are cross-correlated. The orders of the various polynomials are q, q^S, and q^N for moving average polynomials, and p, p^S, and p^N for full autoregressive polynomials (including the differencing operators).

It is natural to ask whether θ^S, ϕ^S, θ^N, and ϕ^N can be computed from a knowledge of θ and ϕ; such a procedure is called a component decomposition. When no hysteresis is present, Hillmer and Tiao (1982) describe a procedure based on a partial fraction decomposition of $\theta(z)\theta(\bar{z})$; also see the discussion in Bell and Martin (2004). In the case of full hysteresis, i.e. the innovation sequences are fully correlated, the component decomposition follows from a partial fraction decomposition of $\theta(z)$, as discussed in Brewer, Hagan, and Perazzelli (1975) and Brewer (1979) – also see Proietti (1995) for the seasonal case. More specifically, if $b_t = b a_t$ and $c_t = c a_t$ for some constants b and c, then one solves

$$\theta(z) = b\varphi^N(z)\theta^S(z) + c\varphi^S(z)\theta^N(z)$$

for b, c and θ^S, θ^N, using partial fractions. Apart from these two extreme cases – of either no cross-correlation or full cross-correlation of latent components – there is no published algorithm for obtaining the component decomposition, and this seems to be a challenging mathematical problem (see further discussion in Section 4).

Supposing for now that the component models are known – either by a decomposition algorithm or by a structural estimation of the model (Section 4) – we proceed to describe the hysteretic filter using the ARIMA model notation. First, let $z = e^{-i\lambda}$ for any frequency $\lambda \in [-\pi, \pi]$. The cross-covariance function of the innovation $\{c_t\}$ relative to $\{b_t\}$ is defined via

$$\rho_h = \mathbb{E}[c_{t+h} b_t], \quad h \in \mathbb{Z}$$

which has cross-covariance generating function $\rho(z)\sigma_b\sigma_c$; it follows from the Cauchy–Schwarz inequality that $|\rho(z)| \leq 1$. Then we have:

$$\begin{aligned}
f_U(\lambda) &= \frac{\theta^S(z)\theta^S(\bar{z})}{\phi^S(z)\phi^S(\bar{z})}\sigma_b^2 \\
f_V(\lambda) &= \frac{\theta^N(z)\theta^N(\bar{z})}{\phi^N(z)\phi^N(\bar{z})}\sigma_c^2 \\
f_{UV}(\lambda) &= \frac{\theta^S(z)\theta^N(\bar{z})}{\phi^S(z)\phi^N(\bar{z})}\rho(\bar{z})\sigma_b\sigma_c \\
f_{VU}(\lambda) &= \frac{\theta^N(z)\theta^S(\bar{z})}{\phi^N(z)\phi^S(\bar{z})}\rho(z)\sigma_b\sigma_c.
\end{aligned} \qquad [7]$$

The optimal hysteretic filter arising from application of Theorem 2 will be denoted by Ψ_H, and clearly depends on the function ρ:

$$\Psi_H(z) = \frac{\theta^S(z)\theta^S(\bar{z})\varphi^N(z)\varphi^N(\bar{z})\sigma_b^2 + \theta^S(z)\theta^N(\bar{z})\varphi^N(z)\varphi^S(\bar{z})\rho(\bar{z})\sigma_b\sigma_c}{\theta(z)\theta(\bar{z})\sigma_a^2}. \qquad [8]$$

This follows from substitution of the relevant spectra into Theorem 2. We observe that both the BN and the WK filters (for bi-infinite processes) are special cases of Theorem 2 where $\rho(z)$ is set equal to one and zero, respectively [see Beveridge and Nelson (1981) and Bell (1984)]; this follows from a further algebraic relation among the hysteretic, BN,[4] and WK filters that is derived below. (The case that $\rho(z)$ is set equal to –1 also produces full correlation between components and yields the negative of the BN filter described below.) In what follows, we describe a relationship between the formulas in eqs. [8]–[10], where the polynomials θ, θ^S, θ^N, φ^N, and φ^S are the same numerically in each formula. The formulas for the BN and WK filters are

$$\Psi_{BN}(z) = \frac{\theta^S(z)\varphi^N(z)\sigma_b}{\theta(z)\sigma_a} \qquad [9]$$

$$\Psi_{WK}(z) = \frac{\theta^S(z)\theta^S(\bar{z})\varphi^N(z)\varphi^N(\bar{z})\sigma_b^2}{\theta(z)\theta(\bar{z})\sigma_a^2}. \qquad [10]$$

Since Ψ_{BN} is a signal extraction filter, we may denote it also by Ψ^S_{BN} when there is the need to distinguish it from the Beveridge–Nelson noise extraction filter given by

$$\Psi^N_{BN}(z) = \frac{\theta^N(z)\varphi^S(z)\sigma_c}{\theta(z)\sigma_a},$$

which is in turn equal to the noise hysteretic filter when $\rho \equiv 1$. Then we have

$$\begin{aligned}\Psi_H(z) &= \Psi^S_{BN}(z)\left[\frac{\theta^S(\bar{z})\varphi^N(\bar{z})\sigma_b + \theta^N(\bar{z})\varphi^S(\bar{z})\rho(\bar{z})\sigma_c}{\theta(\bar{z})\sigma_a}\right] \\ &= \Psi^S_{BN}(z)\left[\Psi^S_{BN}(\bar{z}) + \rho(\bar{z})\Psi^N_{BN}(\bar{z})\right] \\ &= \Psi_{WK}(z) + \rho(\bar{z})\Psi^S_{BN}(z)\Psi^N_{BN}(\bar{z}).\end{aligned} \qquad [11]$$

This derivation uses several facts. The first equality arises from splitting the rational function in eq. [8] into polynomials involving z on the left and $\bar{z}$ on the right. The second equality utilizes the conjugate of the BN noise filter. The last equality is a simple rearrangement of terms, recognizing that $\Psi_{WK}(z)$ is the squared modulus of $\Psi^S_{BN}(z)$.

4 See the Appendix for the derivation of the BN filter.

Let us interpret this interesting relationship. When $\rho \equiv 0$, eq. [11] reduces to $\Psi_H = \Psi_{WK}$, which is true because when no hysteresis is present, the optimal signal extraction filter is given by the WK. But when $\rho \equiv 1$, full hysteresis holds and $\Psi^N_{BN} = 1 - \Psi^S_{BN}$, so that the right-hand side of eq. [11] produces

$$\Psi_{WK}(z) + \Psi^S_{BN}(z)(1 - \Psi^S_{BN}(\overline{z})) = \Psi^S_{BN}(z).$$

Thus, the hysteretic filter is equal to the BN filter in this case. More generally, as in the work of Proietti (2006), the innovation correlation function $\rho(\overline{z}) = \rho$ might be constant in $[-1, 1]$, and then the hysteretic filter is just a weighted combination of the familiar WK and BN (signal and noise) filters. When $\rho(\overline{z})$ is non-constant, additional lag and amplitude effects are superimposed on the constituent filters. In summary, setting ρ identically to unity or zero, respectively, in the hysteretic filter (eq. [11]) will yield either the BN or WK filter.

The actual implementation of the finite sample formulas is straightforward once the spectra and cross-spectra are known. To obtain these, one needs to compute the fitted components models that follow eq. [6]. How is this to be accomplished in practice? A current approach in the literature on hysteresis uses structural models with correlated innovations and computes the log Gaussian likelihood using state space algorithms; see Proietti (2006). In Section 4, we provide details on a structural hysteretic model that generalizes the work of Proietti (2006) to seasonal time series, providing a full illustration of how to fit component models that follow eq. [6].

It is also possible for the correlation ρ to arise from the model itself, as in the work of Harvey and Trimbur (2007). In that paper an underlying continuous time model is stipulated for trend and noise components, and the way in which sampled signal and noise are defined implies a hysteretic structure, with correlation derivable from other parameters of the model. Another natural way in which cross-component correlation may arise is through the signal extraction estimates themselves, even from a model that utilizes an assumption of zero hysteresis. It is well-known that the WK approach results in signal and noise extractions that are cross-correlated. The cross-covariance generating function induced by the signal and noise extractions has different behavior, depending on whether underlying hysteresis is present. When the components are orthogonal, the cross-covariance generating function converges to zero as the lag increases, whereas this does not happen when the series is nonstationary and hysteresis is present.

Note that for long time series (say more than 360 observations) the matrix approach is still quite fast in practice, since the required inversion of non-Toeplitz matrices of dimension equal to sample size is very fast relative to model fitting. (Signal extraction requires a few matrix inversions, whereas optimizing a likelihood requires such an inversion during each evaluation step; state space methods that utilize the Kalman recursions can be used to reduce computation time.) The extra time required over a state space approach is operationally irrelevant for series of moderate length, and in any event may be warranted if one is interested in the full covariance matrix of the error process. So in summary the procedure is:

1. Obtain fitted component models that follow eq. [6];
2. Compute the autocovariance and cross-covariance functions[5] corresponding to f_U, f_V, f_{UV}, and f_{VU} given in eq. [7];
3. Use these quantities to compute P, M, and F in Theorem 1.

We next illustrate the frf of hysteretic filters, adopting the trend–cycle paradigm studied in Proietti (2006). In particular, the signal is a smooth trend given by the model $(1-B)^2 S_t = b_t$, whereas the noise is a stochastic cycle given by $\Phi(B)N_t = c_t$, and $\Phi(B) = 1 - 2\kappa\cos(\omega)B + \kappa^2 B^2$. The persistence of the cycle is determined by $\kappa \in (0,1)$, while the chief frequency is governed by $\omega \in (0, 2\pi)$. As in Section 4, the correlation between the white noise sequences $\{b_t\}$ and $\{c_t\}$ is $\rho \in [-1,1]$, and they each have variance σ_b^2 and σ_c^2, respectively.

The signal extraction problem is identical with trend estimation, whereas the noise extraction problem corresponds to cycle estimation. A third component of white noise could also be included in this simple process, but this somewhat clouds intuition and is therefore omitted here (we consider three component models in the next section). We calculate the frfs for the hysteretic filter Ψ_H, for both trend and cycle estimation, and plot the resulting components: real and imaginary portions (the frf is complex in general), squared gain (labeled just as gain), and phase delay. See Findley and Martin (2006) for definitions of these frequency domain functions.

The parameters of the cyclical model are $\kappa = .8$ and $\omega = \pi/60$– for a cycle of period roughly 5 years – and $\sigma_b = \sigma_c = 1$. For these parameters, we compute the hysteretic, BN, and WK filters as described above. That is, Ψ_H is computed via eq. [8] using $\rho = \pm .5$, whereas Ψ_{BN} and Ψ_{WK} are computed using eqs. [9] and [10] respectively, essentially ignoring the true values of ρ. However, Ψ_{BN} and Ψ_{WK} still depend implicitly on ρ through eq. [5], because $\theta(z)$ appearing in eqs. [9]

5 Standard algorithms – in R for instance – produce the autocovariances for f_U and f_V; a little more work is required for the cross-covariances, but similar principles are in play.

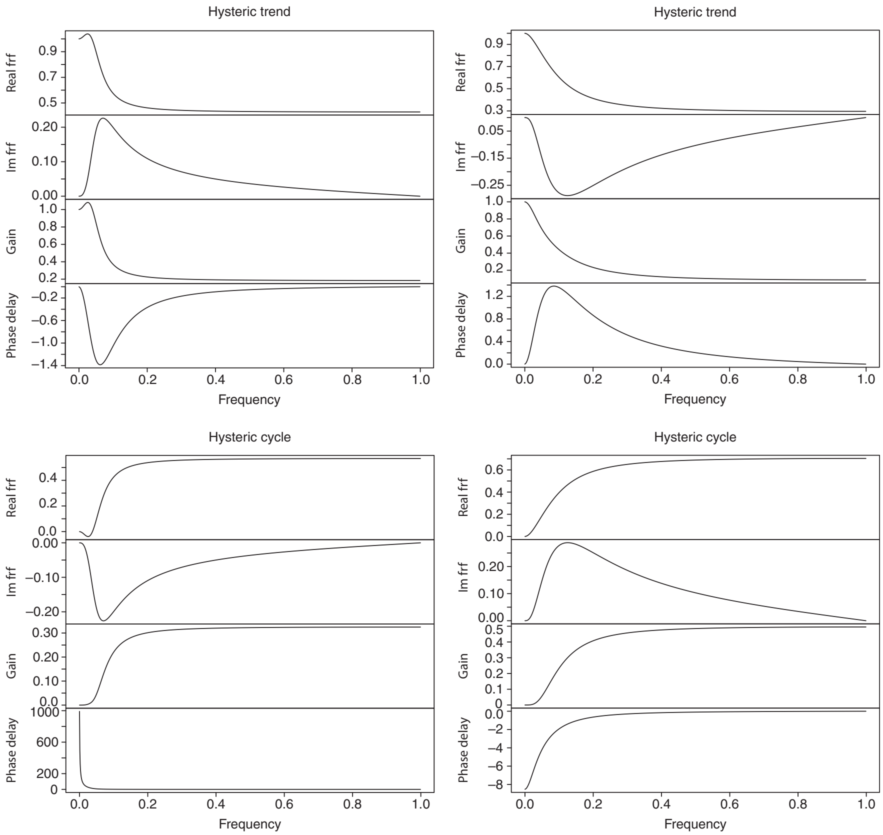

Figure 1: frfs for hysteretic trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part frf, squared gain, and phase delay, moving from top to bottom

and [10] is calculated from f_W, which in turn depends upon the true value of ρ. We have plots for $\rho = \pm .5$: see Figures 1–4.

The squared gain and phase delay are probably of chief interest. For the hysteretic trend in Figure 1, the low-pass shape of the squared gain function is familiar, although in the $\rho = 0.5$ case there is a "nose" in the cycle-band reminiscent of a concurrent filter. For positive cross-correlation, there is phase advance at cyclical frequencies, whereas the opposite effect – phase delay in the cycle band – is evident when $\rho = -0.5$. The cycle filter squared gain functions are as expected, and there is considerable phase advance at low frequencies when $\rho = -0.5$.

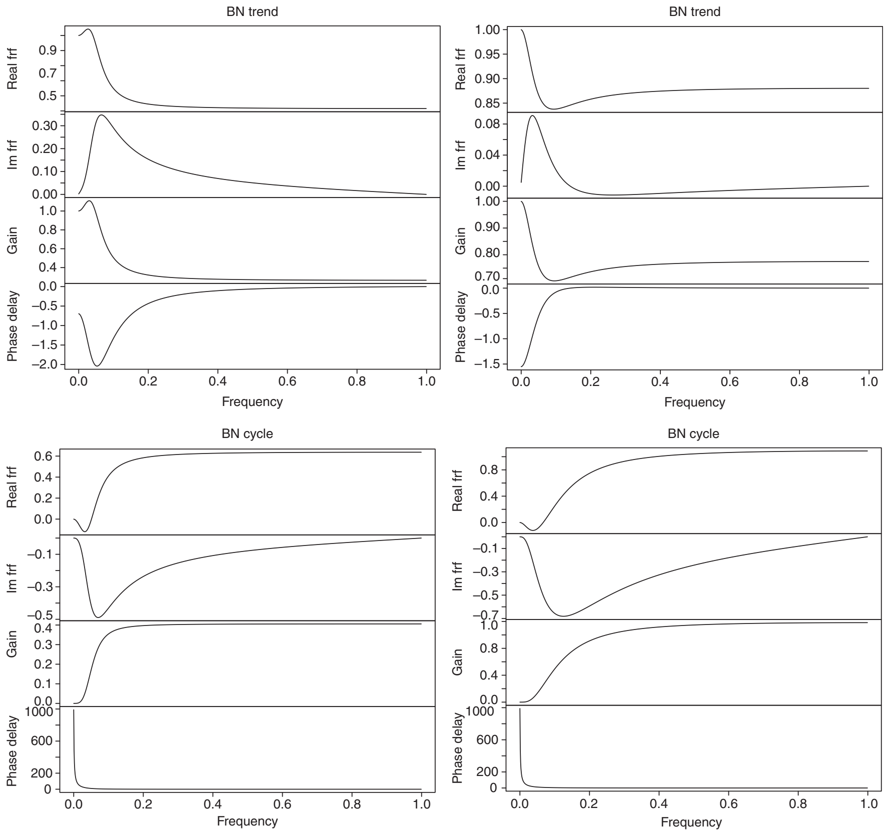

Figure 2: frfs for BN trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part frf, squared gain, and phase delay, moving from top to bottom

The squared gain functions for the BN trend filter in Figure 2 are qualitatively similar to the hysteretic case. The phase delay functions for the BN cycle filter is not well-defined at frequency zero, resulting in non-informative explosive behavior. For the WK case in Figure 3, there is no phase delay (since the filters are symmetric). The low-pass shape for the trend and cycle filters is even more pronounced in the WK case, but qualitatively the impact of $\rho = 0.5$ versus $\rho = -0.5$ is the same.

Finally, we track the relations between the hysteretic, BN, and WK filters in Figure 4. This is the visual display corresponding to eq. [11], where $\rho = \pm 0.5$. We don't examine squared gain or phase delay comparisons, since eq. [11] does not apply to these nonlinear functions of the frf. The trend (real part) frfs for

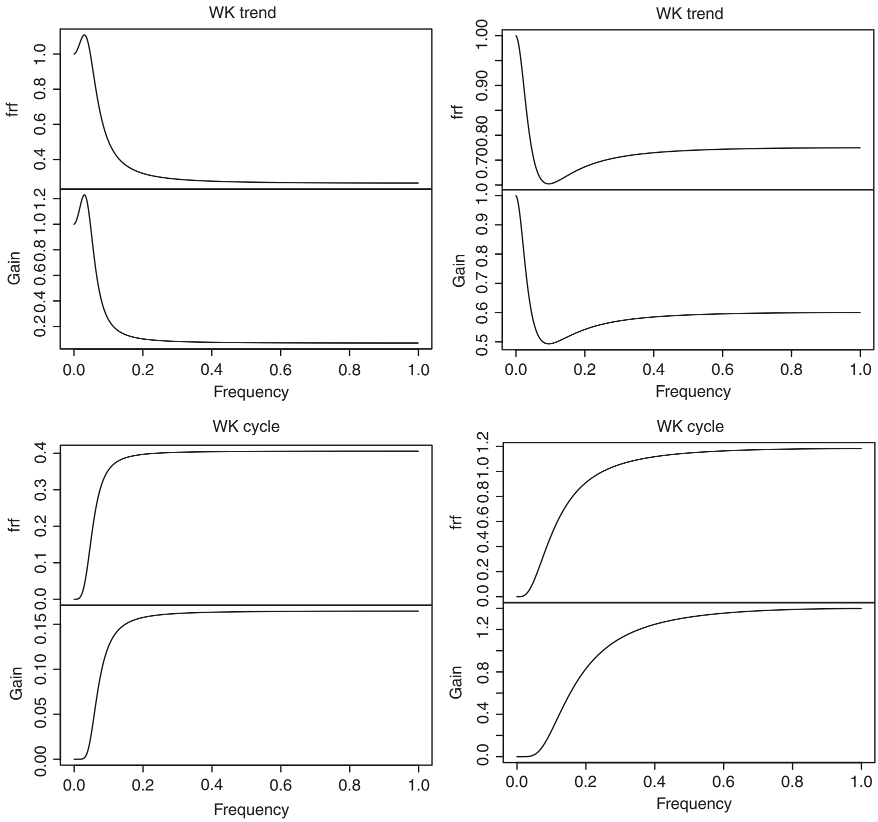

Figure 3: frfs for WK trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part frf, squared gain, and phase delay, moving from top to bottom

$\rho = -0.5$ show some dramatic differences, with the hysteretic filter providing much more smoothing than the BN and WK filters. For the cycle filters, there is an interesting negative dip in hysteretic and BN filters – associated with their asymmetry – at the low frequencies.

4 Estimation of structural hysteretic models

In order to utilize the formulas of Theorem 1, it is necessary to have the autocovariance and cross-covariance functions for the components. This section

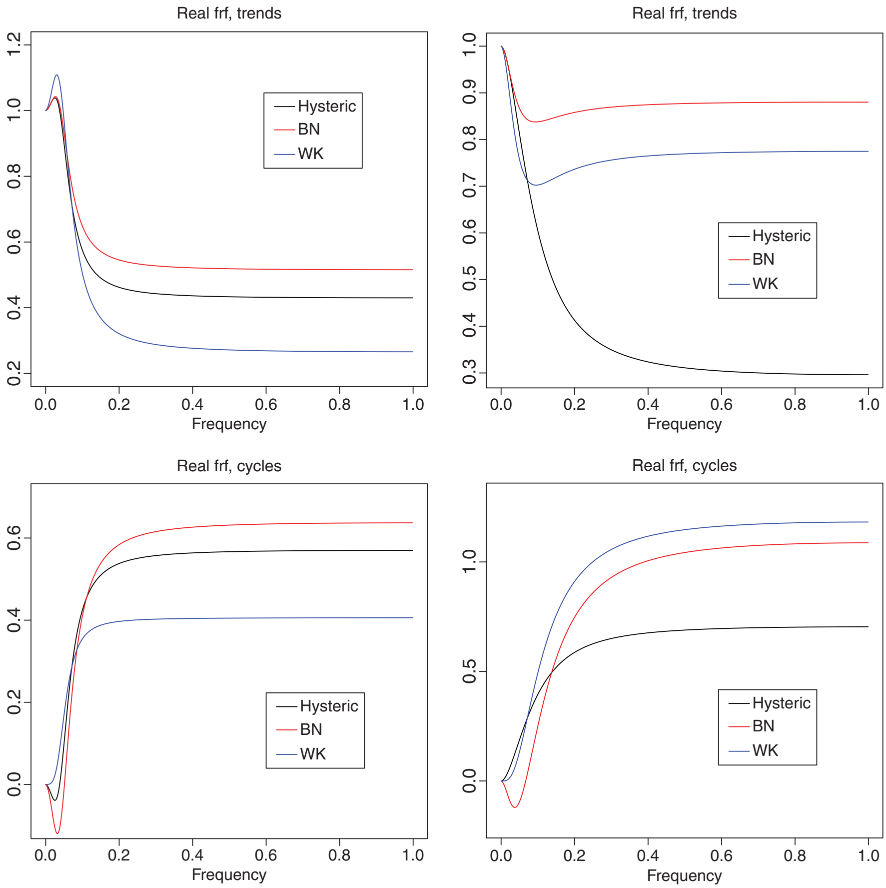

Figure 4: Real part of frfs for trend and cycle filters, for hysteresis, BN, and WK approaches. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -0.5$. Bottom panels are for cycle filtering, top panels are for trend filtering

discusses the general approach to this problem and then considers the identifiability and estimation of a particular basic structural model.

4.1 Latent trend, seasonal, and irregular components

The two most popular model-based approaches in the econometrics literature for obtaining models for latent components are the structural and decomposition methods. The former postulates models for the unobserved signal and noise –

and this includes cross-correlation structure – so that for any choice of parameters the autocovariances for the aggregate process are calculated via eq. [4]. In this way, Gaussian maximum likelihood estimation yields the component models, and the signal extraction formulas can then be applied. (Alternatively, one can utilize the Whittle likelihood by using the spectral representation of the model.) In contrast, the decomposition approach begins with a posited model for the aggregate process and obtains models for the components using algebra; see Hillmer and Tiao (1982) for the orthogonal case and Brewer (1979) for the collinear case.

Much of the current work on hysteresis (Proietti 2006) adopts the structural approach within the context of Harvey's basic structural models, modeling the cross-correlation via cross-correlated white noise innovation sequences for signal and noise. The model estimation and signal extraction computations (corresponding to the matrix formulas of Theorem 1) are efficiently carried out through state space algorithms. The results of Theorem 2 can be used to understand the frequency domain properties of these signal extraction filters.

We now describe the methodology for fitting a structural hysteretic model; we adopt the approach of direct specification of the components in order to simplify identification issues. A typical application to economic time series will involve a model with seasonal, trend, and irregular components. So we posit the existence of a seasonal component P (typically an S is used, but to avoid confusion with the signal notation, we use P for periodic), a trend component T, and a (white noise) irregular component I such that $Y = P + T + I$. The seasonal differencing and trend differencing operators are denoted by $\delta^P(z)$ and $\delta^T(z)$, respectively, such that their application to $\{P_t\}$ and $\{T_t\}$, respectively, yield weakly stationary series. Hence we have by assumption

$$\varphi^Y(B)Y_t = \Theta^Y(B)a_t$$

$$\varphi^P(B)P_t = \Theta^P(B)b_t$$

$$\varphi^T(B)T_t = \Theta^T(B)c_t$$

for innovation sequences $\{a_t\}$, $\{b_t\}$, and $\{c_t\}$, which along with $\{I_t\}$, a white noise. We use Θ to denote moving average polynomials, whereas φ denotes a combined autoregressive and unit root differencing polynomial. As mentioned above, we use δ for differencing polynomials; if we wish to refer to the non-unit root factors of φ, i.e. the pure autoregressive portions, then we use the symbol ϕ. These conventions apply to aggregate Y, trend T, and seasonal P. It is convenient in the following treatment to assume that $\varphi^Y = \varphi^T\varphi^P$.

We assume that φ^P and φ^T are relatively prime; this only constitutes a decision about how the signals are conceived. Allowing for correlation between b_t, c_t, and I_t– but for simplicity we suppose that cross-correlation only occurs at lag zero – produces the following analog of eq. [5]:

$$\frac{\Theta^Y(z)\Theta^Y(\bar{z})}{\varphi^Y(z)\varphi^Y(\bar{z})}\sigma_a^2 = \frac{\Theta^P(z)\Theta^P(\bar{z})}{\varphi^P(z)\varphi^P(\bar{z})}\sigma_b^2 + \frac{\Theta^T(z)\Theta^T(\bar{z})}{\varphi^T(z)\varphi^T(\bar{z})}\sigma_c^2 + \sigma_I^2$$
$$+ \frac{\Theta^P(z)}{\varphi^P(z)}\tau\sigma_b\sigma_I + \frac{\Theta^P(\bar{z})}{\varphi^P(\bar{z})}\tau\sigma_b\sigma_I + \frac{\Theta^T(z)}{\varphi^T(z)}\xi\sigma_c\sigma_I + \frac{\Theta^T(\bar{z})}{\varphi^T(\bar{z})}\xi\sigma_c\sigma_I \quad [12]$$
$$+ \frac{\Theta^T(z)\Theta^P(\bar{z})}{\varphi^T(z)\varphi^P(\bar{z})}\rho\sigma_b\sigma_c + \frac{\Theta^P(z)\Theta^T(\bar{z})}{\varphi^P(z)\varphi^T(\bar{z})}\rho\sigma_b\sigma_c.$$

The correlation between b_t and I_t is denoted by τ, the correlation between c_t and I_t is denoted by ξ, and the correlation between b_t and c_t is denoted by ρ. Clearly τ, ξ, and ρ are bounded by one in magnitude; also it follows – by the injunction that the covariance matrix of b_t, c_t, and I_t be non-negative definite – that $1 - (\rho^2 + \xi^2 + \tau^2) + 2\rho\xi\tau \geq 0$. Parameters satisfying this injunction are said to be admissible. Note that any admissible choice of Θ^T, Θ^P, σ_b, σ_c, τ, ξ, and ρ must also satisfy eq. [12], for some polynomial Θ; it is easy to see that $\delta(B)Y_t$ can be expressed as a linear combination of a tri-variate stationary vector process, and hence the left-hand side of eq. [12] will always be non-negative.

Now eq. [12] tells us how the structural components are related to the aggregate process, even as the terms on the right-hand side sum up to the left-hand side. Passing into time domain via integration against $e^{i\lambda h}$, we obtain a relation of autocovariances. Parameters enter into the covariance quantities, which are in turn aggregated to the whole.

4.2 The basic hysteretic structural model

Here we further illustrate the relation of latent components in eq. [12] through a specific class of structural models. Let $\varphi^P(z) = U(z) = 1 + z + \cdots + z^{11}$ and $\varphi^T(z) = (1 - z)^d$ with $d = 1, 2$. Also set $\Theta^P(z)$ and $\Theta^T(z)$ equal to unity by fiat, so that the only parameters are $\psi = (\sigma_b^2, \sigma_c^2, \sigma_I^2, \rho, \tau, \xi)'$. This corresponds to the Local Level Model or Smooth Trend Model, as well as the basic seasonal model, popularized in Harvey (1989) and utilized in the software STAMP. Here we allow for nonzero correlation between the three innovations.

Then with $W_t = \delta(B)Y_t$, the differenced aggregate process' spectral density is given by

$$
\begin{aligned}
f_W(\lambda) = {} & \sigma_b^2|1-z|^{2d} + \sigma_c^2|U(z)|^2 + \sigma_I^2|1-z|^{2d}|U(z)|^2 \\
& + \tau\sigma_b\sigma_I(U(z)+U(\bar{z}))|1-z|^{2d} + \xi\sigma_c\sigma_I\left((1-z)^d+(1-\bar{z})^d\right)|U(z)|^2 \\
& + \rho\sigma_b\sigma_c\left((1-z)^d U(\bar{z}) + (1-\bar{z})^d U(z)\right).
\end{aligned}
\tag{13}
$$

Estimation is now straightforward in principle, because the exact Gaussian likelihood for the differenced aggregate variables can be easily computed from the above expression – see Section 4.3. However, the model with $d=1$ is actually not identified, whereas the $d=2$ case is indeed identifiable; details are provided in the following discussion. In applications we therefore restrict attention to the $d=2$ case, and henceforth refer to this model [13] as the Basic Hysteretic Structural Model (BHSM).

We give some details on the identifiability of the BHSM. The autocovariance and cross-covariance sequences for each of the six terms in eq. [13] are given as follows. We denote the cross-covariance functions for differenced components as γ_{PT}, γ_{PI}, etc., with an argument z denoting the corresponding generating function. Then eq. [13] can be rewritten as

$$
\begin{aligned}
\gamma_{WW}(z) = {} & \sigma_b^2\gamma_{PP}(z) + \sigma_c^2\gamma_{TT}(z) + \sigma_I^2\gamma_{II}(z) \\
& + \tau\sigma_b\sigma_I(\gamma_{PI}(z)+\gamma_{PI}(\bar{z})) + \xi\sigma_c\sigma_I(\gamma_{TI}(z)+\gamma_{TI}(\bar{z})) \\
& + \rho\sigma_b\sigma_c(\gamma_{PT}(z)+\gamma_{PT}(\bar{z})).
\end{aligned}
$$

This equation involves six separate functions of z, which are actually self-conjugate (and hence real-valued) with symmetric coefficients. Letting $\zeta_j = z^j + z^{-j}$ as a shorthand, these autocovariance and cross-covariance functions are given by

$$\gamma_{PP}(z) = 2 - \zeta_1$$

$$\gamma_{TT}(z) = 12 + 11\zeta_1 + 10\zeta_2 + 9\zeta_3 + \cdots + \zeta_{11}$$

$$\gamma_{II}(z) = 2 - \zeta_{12}$$

$$\gamma_{PI}(z) + \gamma_{PI}(\bar{z}) = 2 - \zeta_1 + \zeta_{11} - \zeta_{12}$$

$$\gamma_{TI}(z) + \gamma_{TI}(\bar{z}) = 2 - \zeta_{12}$$

$$\gamma_{PT}(z) + \gamma_{PT}(\bar{z}) = -\zeta_1 + \zeta_{11}$$

when $d = 1$. Note that $\gamma_{II}(z)$ and $\gamma_{TI}(z) + \gamma_{TI}(\bar{z})$ are identical, essentially because $|1-z|^2 = (1-z) + (1-\bar{z})$. This means that the model is not identifiable; for example, if $\sigma_b^2 = 0$ and $\sigma_c^2 = 1$, then both the setting $\sigma_I^2 = 2, \xi = 0$ and the setting $\sigma_I^2 = 1, \xi = 1$ yield $\sigma_I^2 + \xi \sigma_c \sigma_I = 2$, so that f_{WW} is not an injective function of the parameters. For $d = 2$ the autocovariances and cross-covariances are given by

$$\gamma_{PP}(z) = 6 - 4\zeta_1 + \zeta_2$$

$$\gamma_{TT}(z) = 12 + 11\zeta_1 + 10\zeta_2 + 9\zeta_3 + \cdots + \zeta_{11}$$

$$\gamma_{II}(z) = 4 - 2\zeta_1 + \zeta_{11} - 2\zeta_{12} + \zeta_{13}$$

$$\gamma_{PI}(z) + \gamma_{PI}(\bar{z}) = 6 - 4\zeta_1 + \zeta_2 - \zeta_{10} + 3\zeta_{11} - 3\zeta_{12} + \zeta_{13}$$

$$\gamma_{TI}(z) + \gamma_{TI}(\bar{z}) = -2\zeta_1 + \zeta_{11} + \zeta_{13}$$

$$\gamma_{PT}(z) + \gamma_{PT}(\bar{z}) = -\zeta_1 + \zeta_2 - \zeta_{10} + \zeta_{11}.$$

In this case, the six generating functions are distinct, and the model is identifiable. We briefly discuss identifiability for structural moving average models, but more detail can be found in Morley, Nelson, and Zivot (2003). Suppose the spectrum for an MA(q) process $\{W_t\}$ can be expressed as $f_{WW} = \sum_{j=1}^m \beta_j f_j$ for real numbers $\beta_1, \beta_2, \ldots, \beta_m$(these are either parameters or some function of the parameters), and real-valued functions f_j, all of which have a q-dependent covariance sequence. Transforming the equation via the inverse Fourier transform then relates the $q+1$ nonzero autocovariances of $\{W_t\}$ to the transforms of the f_j, summed against the β_j. Due to the fact that each f_j is real-valued, the inverse Fourier transforms are a symmetric sequence, and there are $q+1$ nonzero corresponding covariances in each case. Writing each of these as a vector $\gamma(f_j)$, the vector of autocovariances for $\{W_t\}$ is equal to $\sum_{j=1}^m \beta_j \gamma(f_j)$, and identifiability of the β_j quantities is equivalent to the linear independence of the $\gamma(f_j)$ vectors.

To see this, let $\beta = [\beta_1, \ldots, \beta_m]'$ and suppose both β and $\tilde{\beta}$ yield the exact same function f_{WW}; then $\sum_{j=1}^m (\beta_j - \tilde{\beta}_j)\gamma(f_j)$ equals the zero vector, and linear independence of the $\gamma(f_j)$ implies that $\beta = \tilde{\beta}$, i.e. identifiability. But linear dependence means that some distinct $\tilde{\beta} = \beta + r$ exists that yields the same f_{WW} as β, making the model non-identifiable. Furthermore, in the case of our BHSM we can uniquely solve for the parameters given the β_j(provided such a solution exists – here $\beta_1, \beta_2, \beta_3 \geq 0$ must hold). Thus, a sufficient condition for the identifiability of the six BHSM parameters is the linear independence of the six covariance vectors given above. Hence, when $d = 1$ the linear dependence becomes obvious, two of the vectors being identical (the matrix made of the six columns vectors actually has rank 4), whereas with $d = 2$ we can simply verify that the matrix of six column vectors has rank 6.

4.3 Estimation of the BHSM

We next discuss parametrization and maximum likelihood estimation of the BHSM. Recalling that $W = \Delta Y$ and letting its covariance matrix be denoted by Γ_W, our scaled log likelihood is

$$W' \Gamma_W^{-1} W + \log |\Gamma_W|.$$

This uses the assumption – common in signal extraction problems in time series analysis – that the initial $d + 11$ values of the sample are uncorrelated with $\{W_t\}$. Evaluation is possible if we can calculate the autocovariances in Γ_W from eq. [13]. Minimizing the above objective function yields the maximum likelihood estimates (MLEs) for the parameter vector ψ, and these will be denoted by $\widehat{\psi}$.

An interesting facet of evaluating the likelihood is guaranteeing that the three innovations have a positive definite correlation matrix. Our approach follows that of Pinheiro and Bates (1996). Define a sequence of "pre-parameters" $\vartheta = (\vartheta_1, \ldots, \vartheta_6)'$ that are unconstrained (i.e. can be any real number) and are mapped into the constrained vector ψ as described below. The mapping guarantees that variances are positive and the joint covariance matrix is positive definite. We preserve the notation of Pinheiro and Bates (1996) for convenience. The idea is to decompose the joint covariance matrix of the three innovation sequences into its Cholesky factors and then do a spherical coordinates transform on the six free variables of the Cholesky factor:

$$\ell_{11} = \exp \vartheta_1 \qquad \ell_{21} = \exp \vartheta_2 \qquad \ell_{31} = \exp \vartheta_3$$

$$\ell_{22} = \pi \frac{\exp \vartheta_4}{1 + \exp \vartheta_4} \quad \ell_{32} = \pi \frac{\exp \vartheta_5}{1 + \exp \vartheta_5} \quad \ell_{33} = \pi \frac{\exp \vartheta_6}{1 + \exp \vartheta_6}.$$

These six new variables are derived through simple exponential and logistic transforms. Of course an arc-tangent function could be utilized for the latter three variables, but the logistic is convenient for differential calculations later. Next, the Cholesky factor L is upper triangular with entries

$$\begin{bmatrix} \ell_{11} & \ell_{21}\cos(\ell_{22}) & \ell_{31}\cos(\ell_{32}) \\ 0 & \ell_{21}\sin(\ell_{22}) & \ell_{31}\sin(\ell_{32})\cos(\ell_{33}) \\ 0 & 0 & \ell_{31}\sin(\ell_{32})\sin(\ell_{33}). \end{bmatrix}$$

Finally, the covariance matrix of the innovations is $L'L$. After some algebra, we discover that

$$\psi_1 = \exp\{2\vartheta_1\} \quad \psi_2 = \exp\{2\vartheta_2\} \quad \psi_3 = \exp\{2\vartheta_3\}$$

$$\psi_4 = \cos(\ell_{22}) \quad \psi_5 = \cos(\ell_{32}) \quad \psi_6 = \cos(\ell_{22})\cos(\ell_{32}) + \sin(\ell_{22})\sin(\ell_{32})\cos(\ell_{33}).$$

One can easily check that the determinant (with $\rho = \psi_4$, $\tau = \psi_5$, $\xi = \psi_6$) equals $\sin^2(\ell_{22})\sin^2(\ell_{32})\cos^2(\ell_{33})$ times $\psi_1\psi_2\psi_3$, and hence is always non-negative.

One can initialize a maximum likelihood estimation routine with $\vartheta_j = 0$ for all j. This, in a sense, places one at the center of the six-dimensional manifold that the parameter vector belongs to. It is easy to obtain the standard errors for the original parametrization. If the MLE for ϑ is asymptotically normal at rate $\sqrt{n}$ and variance matrix V, approximated by the numerical Hessian, then the MLE for ψ has asymptotic variance DVD' with $D_{jk} = \frac{\partial}{\partial\vartheta_k}\psi_j(\vartheta)$, expressed as a function of ϑ and with the MLE $\widehat{\vartheta}$ plugged in. The matrix D follows from calculus. The first five rows make up a diagonal matrix with entries

$$D_{11} = 2\exp\{2\vartheta_1\}$$

$$D_{22} = 2\exp\{2\vartheta_2\}$$

$$D_{33} = 2\exp\{2\vartheta_3\}$$

$$D_{44} = -\sin\left(\frac{\pi\exp\vartheta_4}{1+\exp\vartheta_4}\right)\frac{\pi\exp\vartheta_4}{(1+\exp\vartheta_4)^2}$$

$$D_{55} = -\sin\left(\frac{\pi\exp\vartheta_5}{1+\exp\vartheta_5}\right)\frac{\pi\exp\vartheta_5}{(1+\exp\vartheta_5)^2}.$$

The last row of D has nonzero entries in its last three columns:

$$D_{64} = -\sin\left(\frac{\pi \exp \vartheta_4}{1+\exp \vartheta_4}\right) \frac{\pi \exp \vartheta_4}{(1+\exp \vartheta_4)^2} \cdot \cos\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right)$$
$$+ \cos\left(\frac{\pi \exp \vartheta_4}{1+\exp \vartheta_4}\right) \frac{\pi \exp \vartheta_4}{(1+\exp \vartheta_4)^2} \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right) \cdot \cos\left(\frac{\pi \exp \vartheta_6}{1+\exp \vartheta_6}\right)$$

$$D_{65} = -\cos\left(\frac{\pi \exp \vartheta_4}{1+\exp \vartheta_4}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right) \frac{\pi \exp \vartheta_5}{(1+\exp \vartheta_5)^2}$$
$$+ \sin\left(\frac{\pi \exp \vartheta_4}{1+\exp \vartheta_4}\right) \cdot \cos\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right) \frac{\pi \exp \vartheta_5}{(1+\exp \vartheta_5)^2} \cdot \cos\left(\frac{\pi \exp \vartheta_6}{1+\exp \vartheta_6}\right)$$

$$D_{66} = \sin\left(\frac{\pi \exp \vartheta_4}{1+\exp \vartheta_4}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_6}{1+\exp \vartheta_6}\right) \frac{\pi \exp \vartheta_6}{(1+\exp \vartheta_6)^2}$$

Suppose now that we wish to estimate a constrained structural hysteretic model, where some of the ρ, τ, ξ parameters are forced to be zero. Clearly these generate nested models, and conveniently the nested model is *not* on the boundary of the parameter space of the nesting model. Then the Gaussian Likelihood Ratio (GLR) theory of Taniguchi and Kakizawa (2000) can be applied: we take the difference of the log likelihoods, and the asymptotic distribution is χ^2_r where r is the number of correlations set to zero. Now if one correlation is zero, the determinant of the covariance matrix has the form $1 - (x^2 + y^2)$ times the product of the innovation variances, and hence we can utilize a simple circular coordinate transform. Letting ψ_4 and ψ_5 denote the two nonzero correlations, we have

$$\psi_4 = \frac{\exp \vartheta_4}{1+\exp \vartheta_4} \cos\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right)$$

$$\psi_5 = \frac{\exp \vartheta_4}{1+\exp \vartheta_4} \sin\left(\frac{\pi \exp \vartheta_5}{1+\exp \vartheta_5}\right).$$

But if two correlations are fixed at zero, then just let

$$\psi_4 = \frac{\exp\{\vartheta_4\} - 1}{1+\exp \vartheta_4}.$$

In this manner, the implementation of the BHSM is achieved. R code is available from the first author upon request.

5 Data applications

Our objective is not to decide the question of the extent to which hysteresis exist in economic time series, but rather to show how model estimation and signal extraction may be done when hysteresis is present. To that end we examine the fitting of the BHSM to multiple economic time series. We begin our data analysis with 88 U.S. Census Bureau monthly time series for which the Box–Jenkins airline model was selected by the automatic modeling procedure of TRAMO.[6] We then fit the BHSM to each of these series, looking for cases where the hysteretic model's performance is fairly competitive with that of the airline model. There were six cases where this occurred, and we focus our analysis on two of these six for which the likelihoods were better.

The initial 88 series were selected from a broader class of 146 time series, consisting of 39 foreign trade, 10 retail, 10 housing, and 87 manufacturing time series. Most of these exhibit trading day effects, about half have significant Easter effects, and many have additive and/or level shift outliers. Accounting for transformations and fixed effects, produced a subset of 88 series[7] for which the Box–Jenkins airline model was selected, most of which required a log transformation. Therefore, in fitting the BHSM to these models, we are in essence comparing the hysteretic model against the best SARIMA models obtained for these series.

Series m42110 and x42100 – both of them Foreign Trade series – had BHSM models with likelihoods superior to the airline model. Table 1 summarizes the fits, with standard errors in parentheses. For the airline models, all parameters were significantly different from zero, and the values close to 0.6 for the moving average parameters are not uncommonly encountered for monthly economic series. On the other hand, the variance parameters for the BHSMs are also significantly different from zero and are generally smaller than the innovation variance of the airline models. Actually, while the irregular variance σ_I^2 is comparable in magnitude to the airline innovation variance σ^2, the other variances – especially that for the trend – are much smaller in magnitude. However, these values are compatible with our experience with canonical decompositions of the airline model. For m42110 the three correlation parameters are also clearly significant, but approach full collinearity; for x42100 the uncertainty in ρ and ξ is extremely high, marring their interpretability.

6 This is software developed at the Bank of Spain (Maravall and Caporello 2004).

7 34 foreign trade, 4 retail, 6 housing, and 44 manufacturing.

Table 1: MLEs, likelihoods, and AIC for series m42110 and x42100, for both the airline model and the BHSM

Airline model	**m42110**	**x42100**
Lik	287.60	238.16
AIC	−569.20	−470.31
θ	0.65 (0.046)	0.65 (0.120)
Θ	0.56 (0.053)	0.65 (0.151)
σ^2	0.0062 (0.00052)	0.012 (0.0033)
BHSM	m42110	x42100
Lik	288.35	241.49
AIC	−564.69	−470.97
σ_b^2	0.00028 (0.00008)	0.00066 (0.000196)
σ_c^2	0.0000066 (0.0000027)	0.000022 (0.0000081)
σ_I^2	0.0020 (0.00031)	0.0088 (0.000887)
ρ	−0.993 (0.099)	−0.519 (0.363)
τ	0.995 (0.111)	−1.000 (0.0000033)
ξ	−1.000 (0.0149)	0.519 (0.362)

Note: Standard errors for parameter estimates are in parentheses.

It is clear that for m42110 the BHSM is only marginally superior to the airline model, and from the perspective of Akaike Information Criterion (AIC) is not better at all (note that we cannot use the χ^2 GLR test since the models are not nested). For x42100 the AIC of the BHSM is barely superior to that of the airline model – and given the statistical uncertainty in the likelihoods, it is hard to prefer one over the other. Both BHSMs have some full correlations, indicating redundancies among the innovation sequences.[8] Redundancy in this case means full collinearity of some of the innovations, implying the possibility of a more simplified model. All the estimated coefficients are at least two standard errors away from zero, excepting the ρ and ξ correlations in the BHSM fit of x42100.

For these series, we can assess model goodness-of-fit via computing the model residuals. One only needs the covariance matrix of the differenced

8 This phenomenon was quite common for the BHSM fitted to other series; redundancy in the parameter vector ψ manifested numerically through a final Hessian with some eigenvalues equal to zero, or even a negative number. The latter case indicates a saddle point in the likelihood, where routines such as Nelder–Mead, BFGS, and simulated annealing typically fail. However, m42110 and x42100 did not have this problem, their Hessians being positive definite.

aggregate variables evaluated at the MLEs, which will be denoted $\widehat{\Gamma}_W$. Then the time series residuals are

$$R = \widehat{\Gamma}_W^{-1/2} W,$$

where we can use a matrix square root or a Cholesky factor to define $\widehat{\Gamma}_W^{1/2}$. Note that the χ^2 distribution theory for the Box–Pierce statistic does not apply for the BHSM, since it is not an ARIMA process. The ACF plots of the residuals are given

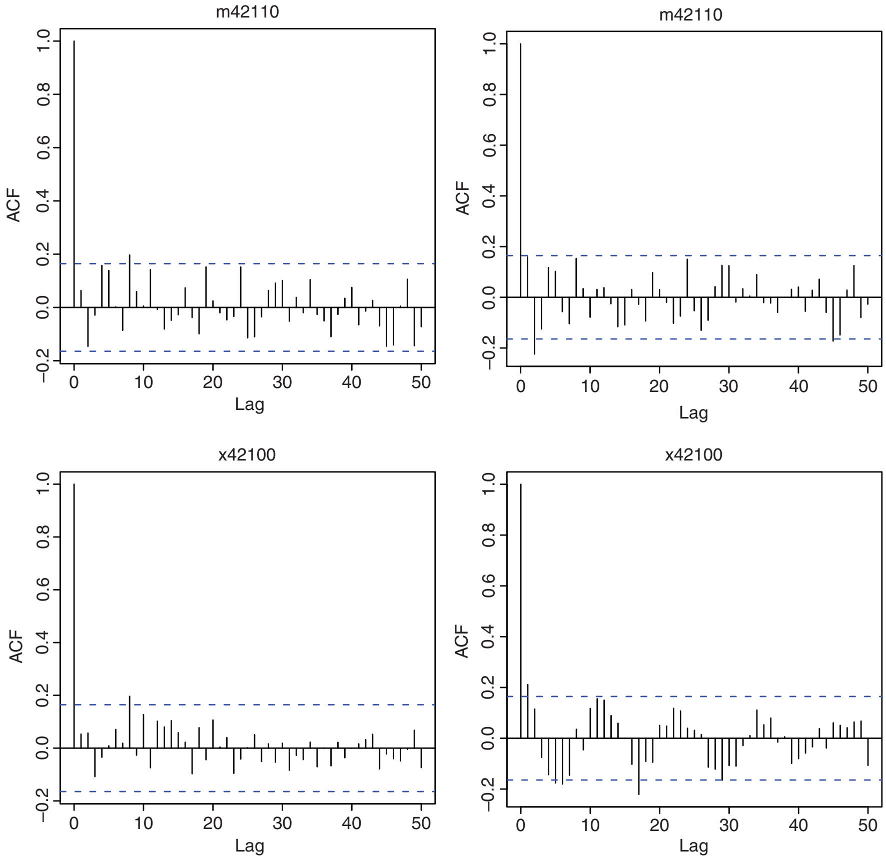

Figure 5: Sample autocorrelation plots for time series residuals. The upper panels correspond to series m42110, while the lower panels correspond to series x42100. The left panels are for the fitted airline model, whereas the right panels are for the fitted BHSM

in Figure 5. Note that some remaining cyclical structure seems to be present in the residuals of the BHSM for x42100, although not many of the correlations are significantly different from zero.

Based upon this analysis, we choose to provide an illustration of the BHSM method on the m42110 series. We reiterate that we are not primarily concerned with empirical validation of the hysteresis hypothesis [see Jäger and Parkinson (1994)], but in how to extract signals of interest when hysteresis is deemed to be present. We next compute extractions of seasonal, trend, nonseasonal, and irregular for these series, along with their standard errors, and compare the same quantities that arise from using the fitted airline model, using an assumption of zero hysteresis. These latter signal extractions are computed using the TRAMO-SEATS software, whereas the BHSM components are extracted using software written in R. Figure 6 displays some of these results, showing the trend, nonseasonal (i.e. seasonal adjustment), and seasonal components, along with time-varying MSEs and the concurrent filters.

The trends arising from the hysteretic and WK methods (with the airline model) are compared against the data in the top left panel of Figure 6. It appears that the trends are slightly more oscillatory in the WK case. In contrast, the TRAMO-SEATS seasonal is slightly more stable than the seasonal extracted using the BHSM model. The seasonal extractions for both approaches are displayed in the top right and middle right panels. In addition, the action of the concurrent filters can be examined in the middle left panel of Figure 6, which displays seasonal adjustment concurrent filters. Both have the characteristic dips – which serve to remove seasonality – but the WK filter places a bit more weight on the current observation. The actual extractions, displayed in the bottom right panel, have the corresponding behavior: the BHSM seasonal adjustment is slightly less erratic, while the WK seasonal adjustment is more timely. Overall, the extractions are quite similar.

The optimal signal extraction matrix F for hysteretic processes need not be centro-symmetric, as is the case with orthogonal components (McElroy 2008), with the result that the time-varying MSEs need not be symmetric with respect to the center point of the sample – this is seen in the bottom left panel of Figure 6. This is an interesting contrast, between the BHSM and airline model's MSE plot. While the lower error in the hysteretic extraction (at least in the center of the sample) might lead one to think it superior, the reader must keep in mind that the actual target signals have different definitions, having disparate stochastic structure. The main point of this comparison is to contrast the symmetry of the WK case with the asymmetry of the hysteretic case.

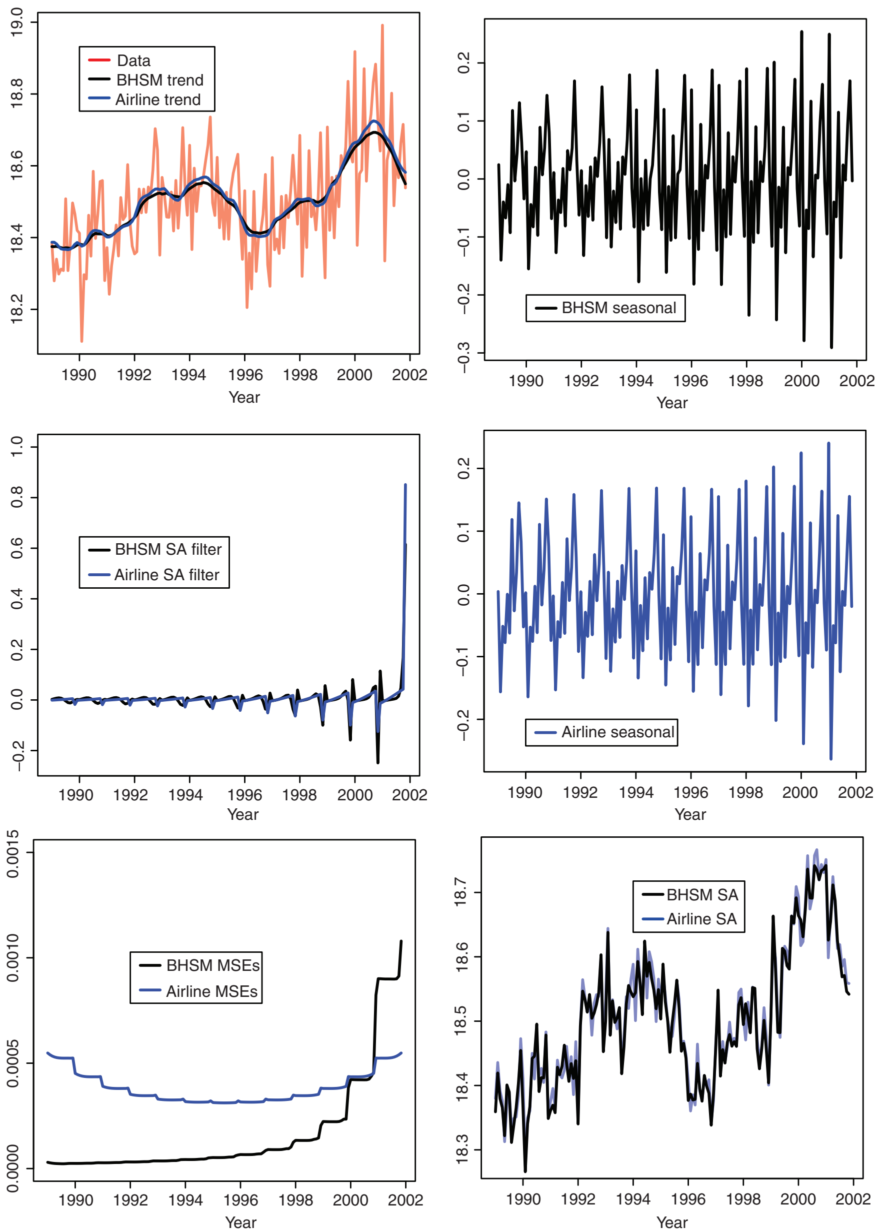

Figure 6: Signal extraction output for m42110, from the BHSM and airline model. The top left panel displays logged data (red) along with the trend extractions (black for BHSM, blue for airline). The upper right panel displays the extracted seasonal for the BHSM, while the middle right panel displays the extracted seasonal for the airline model. The bottom right panel displays extracted seasonal adjustments for both models. The center left panel displays the seasonal adjustment concurrent filter weights for the BHSM and airline models, while the bottom left panel shows the time-varying MSEs for both seasonal adjustments

6 Conclusions

This paper sets out much of the mathematical and statistical theory for hysteretic time series, taking for granted that such processes are of interest to the statistical and econometric community. Indeed, analysis of economic data utilizing hysteretic models has already been going on for several years, primarily via a state space formulation – see Ghysels (1987) and Proietti (2006). In this context, our paper presents several novel contributions: (i) exact MSE optimal signal extraction formulas for both the bi-infinite and the finite sample contexts; (ii) linkages of hysteretic filters to more well-known BN and WK filters in the context of component Auto Regressive Moving Average processes; (iii) frequency domain analysis of the bi-infinite filters, examined for simple trend-cycle processes; (iv) introduction of the BHSM, a straightforward generalization of the classical Basic Structural Model to trend–seasonal–irregular processes having hysteresis; (v) demonstration of the BHSM performance on a Foreign Trade series of the U.S. Census Bureau.

We have not undertaken a justification of the use of hysteretic models in econometrics, or an extensive vetting of the BHSM for empirical analysis. This is beyond the goals of the paper, which are more humble and properly methodological: given that a time series analyst wishes to consider using a hysteretic model for his data, the material in this paper will be crucially important for understanding: (a) how to model it, (b) how to estimate it (both model parameters and the actual signal extraction statistics), and (c) how to think about the properties of hysteretic filters. Whereas some previous literature, alluded to in the introduction, does consider the first point (a), we provide additional material from a fairly broad and general perspective. Hysteretic models may be appropriate for some data, in which case the tools of this paper should prove quite useful for modeling and analysis.

Appendix

Derivation of the BN filter

The original BN filter of Beveridge and Nelson (1981) was applied to nonseasonal time series, yielding a decomposition into permanent and transitory components. This was achieved by supposing the innovations of signal and noise to be identical (rather than orthogonal). We extend this notion of the BN filter to the general scenario outlined in Section 3 (and following) by taking $a_t = b_t = c_t$ in eq. [6]. Then the MSE optimal filters have zero error and are given by decorrelating the aggregate process (via the filter $\varphi(z)/\theta(z)$) followed by re-correlating according to the signal's pattern, namely by the filter $\theta^S(z)/\varphi^S(z)$. Multiplying these filters yields $\theta^S(z)\varphi^N(z)/\theta(z)$. The formula we provide in Section 3 generalizes this treatment slightly to the case of a filter derived under the assumption that the innovations are fully (positively) correlated (i.e. they may have different variances). This leads to the insertion of the factor σ_b/σ_a in the formula for Ψ_{BN}.

By definition, this filter will produce optimal extractions under the full positive correlation hypothesis, and of course the signal and noise extractions aggregate back to the original aggregate variables. This defines the filter; of course it may be applied to data that does not satisfy the hypotheses under which the filter was derived. It is in this sense that we may speak of the hysteretic filter as being a convex combination of WK and BN filters – it is an algebraic fact, involving statistical quantities derived under incompatible stochastic assumptions.

We are not aware of prior references to this generalized BN filter, but are not comfortable claiming that our derivation is novel. This sort of idea, and construction, has close precedents in Morley, Nelson, and Zivot (2003) and Proietti (2006).

We also mention an interesting interpretation of the BN filter in the case of a trend plus noise decomposition, namely that it can be interpreted as the optimal concurrent filter in an orthogonal decomposition. Consider the special case of an $I(1)$ data process $\{Y_t\}$ that satisfies $(1-B)Y_t = \theta(B)a_t$, where the polynomial (or causal power series) $\theta(z)$ is invertible. Then the BN decomposition can be written

$$\theta(B)a_t = \theta(1)a_t + (\theta(B) - \theta(1))a_t,$$

which can be compared to $(1-B)Y_t = (1-B)S_t + (1-B)N_t$. Thus, using the notation of eq. [6] we have $\theta^S(z) = \theta(1)$, a constant, while $\theta^N(z) = (\theta(z) - \theta(1))/(1-z)$ and $\varphi^S(z) = 1-z$ and $\varphi^N(z) = 1$. Note that a single innovation $\{a_t\}$ drives both component processes. The assumed invertibility of $\theta(z)$ ensures that $\theta^N(z)$ has a convergent power series expansion. The BN filter can then be expressed as $\theta(1)/\theta(z)$.

Suppose now that we take the same component model definitions – namely $(1-B)S_t = b_t$, a white noise of variance $\theta^2(1)\sigma_a^2$, and $N_t = \theta^N(B)c_t$, with $\{c_t\}$ another white noise – but instead of having a common innovation drive the components, we assume they are uncorrelated. The concurrent filter in this case, using the formulas of Bell and Martin (2004), is

$$\frac{1-z}{\sigma_a^2\theta(z)}\left[\frac{1-\bar{z}}{\theta(\bar{z})}\,\frac{\sigma_b^2}{(1-z)(1-\bar{z})}\right]_+ = \frac{\sigma_b^2}{\sigma_a^2\,\theta(z)\,\theta(1)},$$

which simplifies to $\theta(1)/\theta(z)$. Here the bracket notation with the plus subscript indicates that in the power series expansion, we only retain the terms corresponding to non-negative powers of z.

Proofs

Proof of Theorem 1. Following the same technique as McElroy (2008), it suffices to show that the error process $\varepsilon = \widehat{S} - S$ is uncorrelated with Y. Using I to denote an identity matrix, the error process is

$$\begin{aligned} \varepsilon &= FN - (I - F)S \\ &= M^{-1}\left[\Delta_N' \Gamma_V^{-1} + P\Gamma_W^{-1} \underline{\Delta}_S\right] V \\ &\quad - M^{-1}\left[\Delta_S' \Gamma_U^{-1} - P\Gamma_W^{-1} \underline{\Delta}_N\right] U. \end{aligned}$$

From Assumption A, it follows that ε is uncorrelated with Y^*. Since Y can be expressed as a linear combination of Y^* and W, as discussed in McElroy (2008), it suffices to show that ε is uncorrelated with W. To that end, we have

$$\begin{aligned} \mathbb{E}[\varepsilon W'] &= M^{-1}\left[\Delta_N' \Gamma_V^{-1} + P\Gamma_W^{-1} \underline{\Delta}_S\right] \left(\Gamma_V \underline{\Delta}_S' + \Gamma_{VU} \underline{\Delta}_N'\right) \\ &\quad - M^{-1}\left[\Delta_S' \Gamma_U^{-1} - P\Gamma_W^{-1} \underline{\Delta}_N\right] \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \\ &= M^{-1}\left[\Delta' + \Delta_N' \Gamma_V^{-1} \Gamma_{VU} \underline{\Delta}_N' + P\Sigma_W^{-1}\left(\underline{\Delta}_S \Gamma_V \underline{\Delta}_S' + \underline{\Delta}_S \Gamma_{VU} \underline{\Delta}_N'\right)\right] \\ &\quad - M^{-1}\left[\Delta' + \Delta_S' \Gamma_U^{-1} \Gamma_{UV} \underline{\Delta}_S' - P\Sigma_W^{-1}\left(\underline{\Delta}_N \Gamma_U \underline{\Delta}_N' + \underline{\Delta}_N \Gamma_{UV} \underline{\Delta}_S'\right)\right] \\ &= M^{-1}(-P + P) = 0, \end{aligned}$$

using eq. [4]. This establishes MSE linear optimality. For the error covariance matrix, we obtain the formula by expanding $\mathbb{E}[\varepsilon\varepsilon']$ and simplifying the algebra.

It may be instructive to offer a constructive derivation of the formula for F. Recall that $U = \Delta_S S$ and $V = \Delta_N N$; the minimum MSE linear estimator of U given Y is the same as the optimal estimator of U given W (by Assumption A), and hence its formula is

$$\mathrm{Cov}(U, W)[\mathrm{Cov}(W, W)]^{-1} W = \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \Gamma_W^{-1} W,$$

which utilizes eq. [3]. By linearity, the above estimate must be equal to Δ_S times the optimal estimate of S given Y. Similarly, we can derive the minimum MSE linear estimator of V given Y via

$$\mathrm{Cov}(V, W)\,[\mathrm{Cov}(W, W)]^{-1} W = \left(\Gamma_V \underline{\Delta}_S' + \Gamma_{VU} \underline{\Delta}_N'\right) \Gamma_W^{-1} W.$$

(This tactic, whereby we directly compute the best linear estimate of S given Y, can be used to derive F as well, but requires considerably more algebra than the approach given here.) Collecting these results, and letting F denote the matrix such that FY is the minimum MSE linear estimator of S given Y, we obtain

$$\Delta_S F = \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \Gamma_W^{-1} \Delta$$

$$\Delta_N F = \Delta_N - \left(\Gamma_V \underline{\Delta}_S' + \Gamma_{VU} \underline{\Delta}_N'\right) \Gamma_W^{-1} \Delta.$$

The second equation utilizes $\Delta_N F = \Delta_N - \Delta_N(I - F)$ and the fact that $I - F$ is the linear optimal estimator of the noise N given Y. Now knowing algebraically the form of both $\Delta_S F$ and $\Delta_N F$ permits us to solve for F, presuming that the intersection of the null spaces of Δ_S and Δ_N is the zero vector; this follows from the assumption that $\delta^S(z)$ and $\delta^N(z)$ are relatively prime. Then multiply $\Delta_S F$ by $\Delta_S' \Gamma_U^{-1}$(this is not a unique choice, e.g. we could just utilize Δ_S' instead and get an alternative expression for the unique F) and $\Delta_N F$ by $\Delta_N' \Gamma_V^{-1}$ and add the result. Then utilizing eq. [2], this yields

$$MF = \Delta' + \Delta_S' \Gamma_U^{-1} \Gamma_{UV} \underline{\Delta}_S' \Gamma_W^{-1} \Delta + \Delta_N' \Gamma_V^{-1} \Delta_N - \Delta' - \Delta_N' \Gamma_V^{-1} \Gamma_{VU} \underline{\Delta}_N' \Gamma_W^{-1} \Delta,$$

which simplifies to $\Delta_N' \Gamma_V^{-1} \Delta_N + P \Gamma_W^{-1} \Delta$. Inverting M then yields the formula for F. □

Proof of Theorem 2. Our strategy is to demonstrate that the filter $\Psi(z)$ produces a signal extraction error process that is orthogonal to the aggregate process, implying MSE linear optimality [cf. Bell (1984)]. It is clear from the given formula that $\delta^N(z)$ can be factored out, leaving $\Psi(z) = \Omega(z)\delta^N(z)$ with

$$\Omega(z) = \frac{f_U(\lambda)\delta^N(\bar{z}) + f_{UV}(\lambda)\delta^S(\bar{z})}{f_W(\lambda)}.$$

Similarly, it is easily verified that $1 - \Psi(z) = \Phi(z)\delta^S(z)$ with

$$\Phi(z) = \frac{f_V(\lambda)\delta^S(\bar{z}) + f_{VU}(\lambda)\delta^N(\bar{z})}{f_W(\lambda)}.$$

Then $\varepsilon_t = \Psi(B)Y_t - S_t = \Omega(B)V_t - \Phi(B)U_t$. This shows that $\{\varepsilon_t\}$ is weakly stationary. The aggregate process $\{Y_t\}$ can be written in terms of a linear combination of d initial values summed with a linear function of the differenced process $\{W_t\}$– see Bell (1984). Hence by Assumption A, it is sufficient to demonstrate that ε_t is uncorrelated with W_{t+h} for any t and h. Now $W_{t+h} = \delta^N(B)U_{t+h} + \delta^S(B)V_{t+h}$, so that

$$\mathbb{E}[\varepsilon_t W_{t+h}] = \Omega(B)\left[\delta^S(B^{-1})\gamma_V(h) + \delta^N(B^{-1})\gamma_{VU}(h)\right] \\ - \Phi(B)\left[\delta^N(B^{-1})\gamma_U(h) + \delta^S(B^{-1})\gamma_{UV}(h)\right],$$

which is independent of t, but holds for all h. Thus, taking the Fourier transform yields

$$\sum_{h\in\mathbb{Z}} z^h\,\mathbb{E}[\varepsilon_t W_{t+h}] = \Omega(z)\left[\delta^S(\bar{z})f_V(\lambda) + \delta^N(\bar{z})f_{VU}(\lambda)\right] \\ - \Phi(z)\left[\delta^N(\bar{z})f_U(\lambda) + \delta^S(\bar{z})f_{UV}(\lambda)\right].$$

Then simple algebra, along with the above formulas for $\Omega(z)$ and $\Phi(z)$, produces $\sum_{h\in\mathbb{Z}} z^h\,\mathbb{E}[\varepsilon_t W_{t+h}] = 0$, and hence $\mathbb{E}[\varepsilon_t W_{t+h}] = 0$ for all h. A second calculation produces

$$\mathbb{E}[\varepsilon_t \varepsilon_{t+h}] = \Omega(B)\Omega(B^{-1})\gamma_V(h) - \Omega(B)\Phi(B^{-1})\gamma_{VU}(h) \\ - \Phi(B)\Omega(B^{-1})\gamma_{UV}(h) + \Phi(B)\Phi(B^{-1})\gamma_U(h).$$

Again, summing against z^h yields

$$f_\varepsilon(\lambda) = \Omega(z)\Omega(\bar{z})f_V(\lambda) - \Omega(z)\Phi(\bar{z})f_{VU}(\lambda) - \Phi(z)\Omega(\bar{z})f_{UV}(\lambda) + \Phi(z)\Phi(\bar{z})f_U(\lambda).$$

Now plugging in for Ω and Φ gives the stated result. □

Alternative calculation of *F*

We mentioned in the proof of Theorem 1 that there is more than one expression for the signal extraction matrix F, although numerically all such formulas are equal. A different approach to the problem, based upon results of Bell and Hillmer (1988), can be developed that has some computational advantages. Let S_* denote the first d_S values of the vector S, conceived of as initial values of the process. Likewise, let N_* denote the first d_N values of the vector N, and Y_* the first d values of the vector Y. As described in Bell and Hillmer (1988), it is always possible to algebraically describe S_* as a linear function of Y_*, U, and V, and likewise for N_*. We review and develop these relationships below.

We use the notation I_n to denote an identity matrix of dimension n. We can relate the signal vector S to its initial values S_* and differenced values U, and similarly for the noise, via the transformations

$$\begin{bmatrix} S_* \\ U \end{bmatrix} = \begin{bmatrix} I_{d_S} & 0 \\ \Delta_S & \end{bmatrix} S \qquad \text{and} \qquad \begin{bmatrix} N_* \\ V \end{bmatrix} = \begin{bmatrix} I_{d_N} & 0 \\ \Delta_N & \end{bmatrix} N.$$

Let us denote these matrices by ∇_S and ∇_N, respectively; they are unit lower triangular and invertible, which allows us to express S directly in terms of S_* and U (and N in terms of N_* and V). Taking only the first d rows of ∇_S^{-1} yields

$$[I_d\ 0]\,S = [I_d\ 0]\,\nabla_S^{-1}\begin{bmatrix} S_* \\ U \end{bmatrix} = \begin{bmatrix} I_{d_S} & 0 \\ A_S & B_S \end{bmatrix}\begin{bmatrix} S_* \\ U \end{bmatrix}.$$

This calculation first uses the fact that – because ∇_S is unit lower triangular with first d_S rows given by $[I_{d_S}\ 0]$– the first d_S rows of ∇_S^{-1} are also given by $[I_{d_S}\ 0]$. The matrices A_S and B_S correspond to the next d_N rows of ∇_S^{-1}. With a similar notation and decomposition for N, and noting that $S + N = Y$, we obtain

$$\begin{aligned} Y_* &= [I_d\ 0]\,S + [I_d\ 0]\,N \\ &= \begin{bmatrix} I_{d_S} & 0 \\ A_S & B_S \end{bmatrix}\begin{bmatrix} S_* \\ U \end{bmatrix} + \begin{bmatrix} I_{d_N} & 0 \\ A_N & B_N \end{bmatrix}\begin{bmatrix} N_* \\ V \end{bmatrix} \\ &= \begin{bmatrix} I_{d_S} & I_{d_N} \\ A_S & A_N \end{bmatrix}\begin{bmatrix} S_* \\ N_* \end{bmatrix} + \begin{bmatrix} 0 \\ B_S \end{bmatrix} U + \begin{bmatrix} 0 \\ B_N \end{bmatrix} V. \end{aligned}$$

The $d \times d$ matrix that multiplies the signal and noise initial values is denoted by $[H_1 H_2]$ in Bell and Hillmer (1988) and is there proved to be invertible. We will denote it by the symbol Ω. Note that its inversion is inexpensive due to its relatively low dimension of d. It now follows that

$$S_* = [I_{d_S}\ 0]\,\Omega^{-1}\left(Y_* - \begin{bmatrix} 0 \\ B_S \end{bmatrix} U - \begin{bmatrix} 0 \\ B_N \end{bmatrix} V\right)$$

$$N_* = [0\ I_{d_N}]\,\Omega^{-1}\left(Y_* - \begin{bmatrix} 0 \\ B_S \end{bmatrix} U - \begin{bmatrix} 0 \\ B_N \end{bmatrix} V\right).$$

These relations between signal and noise initial values are an exact algebraic relation, and a direct implication of the nonstationary signal and noise structure; no stochastic assumptions have been used yet. The relations can be utilized to produce signal extraction estimates as follows. From $S = \nabla_S^{-1}[S_*', U']'$ we deduce that the linear optimal estimate of S can be constructed from estimates of S_* and U, followed by application of ∇_S^{-1}. This latter matrix does not require inversion in general, but rather expresses in matrix notation the notion of recursion. Namely, we can always compute S_t from d_S prior values (in time) together with U_t, i.e. $S_t = -\sum_{j=1}^{d_S} \delta_j^S/\delta_0^S S_{t-j} + U_t$. In the proof of Theorem 1, we discussed the optimal linear estimates of U given Y, and estimates of V given Y, which we shall denote by $\widehat{U}$ and $\widehat{V}$. Then the optimal linear estimates of the signal and noise initial values are just

$$\widehat{S_*} = [I_{d_S}\ 0]\,\Omega^{-1}\left(Y_* - \begin{bmatrix} 0 \\ B_S \end{bmatrix}\widehat{U} - \begin{bmatrix} 0 \\ B_N \end{bmatrix}\widehat{V}\right)$$

$$\widehat{N_*} = [0\ I_{d_N}]\,\Omega^{-1}\left(Y_* - \begin{bmatrix} 0 \\ B_S \end{bmatrix}\widehat{U} - \begin{bmatrix} 0 \\ B_N \end{bmatrix}\widehat{V}\right).$$

It is easy to check, using Assumption A, that the error $S_* - \widehat{S_*}$ is orthogonal to Y. The algorithm is to first compute $\widehat{U}$ and $\widehat{V}$ via $\Delta_S F Y$ and $\Delta_N (I - F) Y$, i.e.

$$\widehat{U} = \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \Gamma_W^{-1} W$$

$$\widehat{V} = \left(\Gamma_V \underline{\Delta}_S' + \Gamma_{VU} \underline{\Delta}_N'\right) \Gamma_W^{-1} W.$$

Then the formula for $\widehat{S_*}$ is utilized, and finally $\widehat{S}$ is obtained recursively (and $\widehat{N} = Y - \widehat{S}$). Such an algorithm can avoid the inversion of large matrices, excepting the work involved in inverting Γ_W; however, this is a Toeplitz matrix, and hence the innovations algorithm of Brockwell and Davis (1991) can be utilized.

Some readers may find it illuminating to derive F directly from $\mathrm{Cov}(S, Y)\ \mathrm{Cov}(Y, Y)^{-1}$, which we now derive, utilizing the above expressions. Let ∇ denote the differencing matrix with upper rows $[1_d\ 0]$ and bottom rows Δ. Because W is orthogonal to Y_*, we find that

$$\mathrm{Cov}(Y, Y) = \nabla^{-1} \begin{bmatrix} \mathrm{Cov}(Y_*, Y_*) & 0 \\ 0 & \Gamma_W \end{bmatrix} \nabla^{\dagger}.$$

Moreover, U is orthogonal to Y_*, so that

$$\mathrm{Cov}(S, Y) = \nabla_S^{-1} \begin{bmatrix} \mathrm{Cov}(S_*, Y_*) & \mathrm{Cov}(S_*, W) \\ 0 & \mathrm{Cov}(U, W) \end{bmatrix} \nabla^{\dagger}.$$

Therefore, we obtain

$$\begin{aligned} \mathrm{Cov}(S, Y)\ \mathrm{Cov}(Y, Y)^{-1} &= \nabla_S^{-1} \begin{bmatrix} \mathrm{Cov}(S_*, Y_*) & \mathrm{Cov}(S_*, W) \\ 0 & \Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S' \end{bmatrix} \begin{bmatrix} \mathrm{Cov}(Y_*, Y_*)^{-1},\ 0 \\ \Gamma_W^{-1} \Delta \end{bmatrix} \\ &= \nabla_S^{-1} \begin{bmatrix} \mathrm{Cov}(S_*, Y_*) \left[\mathrm{Cov}(Y_*, Y_*)^{-1},\ 0\right] + \mathrm{Cov}(S_*, W) \Gamma_W^{-1} \Delta \\ \left(\Gamma_U \underline{\Delta}_N' + \Gamma_{UV} \underline{\Delta}_S'\right) \Gamma_W^{-1} \Delta \end{bmatrix}. \end{aligned}$$

Now utilizing the expression of S_* above, written in terms of Y_*, U, and V, it follows that

$$\mathrm{Cov}(S_*, Y_*) = [I_{d_S}\ 0]\,\Omega^{-1}\,\mathrm{Cov}(Y_*, Y_*)$$

$$\mathrm{Cov}(S_*, W) = -[I_{d_S}\ 0]\,\Omega^{-1}\left(\begin{bmatrix}0\\B_S\end{bmatrix}\left[\Gamma_U\,\underline{\Delta}_N' + \Gamma_{UV}\,\underline{\Delta}_S'\right] + \begin{bmatrix}0\\B_N\end{bmatrix}\left[\Gamma_V\,\underline{\Delta}_S' + \Gamma_{VU}\,\underline{\Delta}_N'\right]\right).$$

As a result,

$$\begin{aligned}&\mathrm{Cov}(S,Y)\ \mathrm{Cov}(Y,Y)^{-1}\\ &= \nabla_S^{-1}\begin{bmatrix}[I_{d_S}\ 0]\,\Omega^{-1}\left([I_d\ 0] - \left(\begin{bmatrix}0\\B_S\end{bmatrix}\left[\Gamma_U\,\underline{\Delta}_N' + \Gamma_{UV}\,\underline{\Delta}_S'\right] + \begin{bmatrix}0\\B_N\end{bmatrix}\left[\Gamma_V\,\underline{\Delta}_S' + \Gamma_{VU}\,\underline{\Delta}_N'\right]\right)\Gamma_W^{-1}\Delta\right)\\ \left(\Gamma_U\underline{\Delta}_N' + \Gamma_{UV}\underline{\Delta}_S'\right)\Gamma_W^{-1}\Delta\end{bmatrix}\end{aligned}$$

This is the same formula for F as given above.

References

Bell, W. 1984. "Signal Extraction for Nonstationary Time Series." *The Annals of Statistics* 12:646–64.

Bell, W. 2004. "On RegCOMPONENT Time Series Models and Their Applications." In *State Space and Unobserved Component Models: Theory and Applications*, edited by A. C. Harvey, S. J. Koopman, and N. Shephard. Cambridge, UK: Cambridge University Press.

Bell, W., and S. Hillmer. 1984. "Issues Involved with the Seasonal Adjustment of Economic Time Series." *Journal of Business and Economics Statistics* 2:291–320.

Bell, W., and S. Hillmer. 1988. "A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models." Technical Report, U.S. Census Bureau.

Bell, W., and S. Hillmer. 1991. "Initializing the Kalman Filter for Nonstationary Time Series Models." *Journal of Time Series Analysis* 12:283–300.

Bell, W., and D. Martin. 2004. "Computation of Asymmetric Signal Extraction Filters and Mean Squared Error for ARIMA Component Models." *Journal of Time Series Analysis* 25:603–25.

Beveridge, S., and C. Nelson. 1981. "A New Approach to the Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to the Measurement of the 'Business Cycle'." *Journal of Monetary Economics* 7:151–74.

Brewer, K. 1979. "Seasonal Adjustment of ARIMA Series." *Économie Appliquée* 1:7–22.

Brewer, K., P. Hagan, and P. Perazzelli. 1975. "Seasonal Adjustment Using Box–Jenkins Models." *Bulletin of the International Statistical Institute, Proceedings of the 40th Session* 31:130–36.

Brockwell, P., and R. Davis. 1991. *Time Series: Theory and Methods*. New York: Springer.

Findley, D., and D. Martin. 2006. "Frequency Domain Analyses of SEATS and X-12-ARIMA Seasonal Adjustment Filters for Short and Moderate-Length Time Series." *Journal of Official Statistics* 22:1–34.

Gersch, W., and G. Kitagawa. 1983. "The Prediction of Time Series with Trends and Seasonalities." *Journal of Business and Economics Statistics* 1:253–64.

Ghysels, E. 1987. "Seasonal Extraction in the Presence of Feedback." *Journal of Business and Economics Statistics* 5:191–94.

Harvey, A. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A., and A. Jäger. 1993. "Detrending, Stylized Facts and the Business Cycle." *Journal of Applied Econometrics* 8:231–47.

Harvey, A., and T. Trimbur. 2007. “Trend Estimation, Signal-Noise Ratios, and the Frequency of Observations.” In *Growth and Cycle in the Eurozone*, edited by G. L. Mazzi and G. Savio. Basingstoke, UK: Palgrave MacMillan.

Hillmer, S., and G. Tiao. 1982. “An ARIMA-Model-Based Approach to Seasonal Adjustment.” *Journal of the American Statistical Association* 77:63–70.

Hyndman, R., A. Koehler, R. Snyder, and S. Grose. 2002. “A State Space Framework for Automatic Forecasting Using Exponential Smoothing Methods.” *International Journal of Forecasting* 18:439–54.

Jäger, A., and M. Parkinson. 1994. “Some Evidence on Hysteresis in Unemployment Rates.” *European Economic Review* 38:329–42.

Kaiser, R., and A. Maravall. 2005. “Combining Filter Design with Model-Based Filtering: An Application to Business-Cycle Estimation.” *International Journal of Forecasting* 21:691–710.

Maravall, A. 1995. “Unobserved Components in Economic Time Series.” In *The Handbook of Applied Econometrics*, edited by H. Pesaran and M. Wickens, Chap. 1, 12–72. Oxford: Basil Blackwell.

Maravall, A., and G. Caporello. 2004. “Program TSW: Revised Reference Manual.” Working Paper 2004, Research Department, Bank of Spain.

McElroy, T. 2008. “Matrix Formulas for Nonstationary ARIMA Signal Extraction.” *Econometric Theory* 24:1–22.

McElroy, T. 2012. “An Alternative Model-Based Seasonal Adjustment That Reduces Over-Adjustment.” *Taiwan Economic Forecast and Policy* 43:33–70.

McElroy, T., and A. Sutcliffe. 2006. “An Iterated Parametric Approach to Nonstationary Signal Extraction.” *Computational Statistics and Data Analysis, Special Issue on Signal Extraction* 50:2206–31.

Morley, J., C. Nelson, and E. Zivot. 2003. “Why Are Beveridge–Nelson and Unobserved-Component Decompositions of GDP so Different?” *Review of Economics and Statistics* 85:235–43.

Oh, K., E. Zivot, and D. Creal. 2008. “The Relationship between the Beveridge–Nelson Decomposition and Other Permanent-Transitory Decompositions That Are Popular in Economics.” *Journal of Econometrics* 146:207–19.

Ord, J., A. Koehler, and R. Snyder. 1997. “Estimation and Prediction for a Class of Dynamic Nonlinear Statistical Models.” *Journal of the American Statistical Association* 92:1621–29.

Pinheiro, J., and D. Bates. 1996. “Unconstrained Parameterizations for Variance-Covariance Matrices.” *Statistics and Computing* 6:289–96.

Proietti, T. 1995. “The Beveridge–Nelson Decomposition: Properties and Extensions.” *Journal of the Italian Statistical Society* 4:101–24.

Proietti, T. 2006. “Trend-Cycle Decompositions with Correlated Components.” *Econometric Reviews* 25:61–84.

R Development Core Team. 2012. “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria.

Snyder, R. 1985. “Recursive Estimation of Dynamic Linear Models.” *Journal of the Royal Statistical Society, B* 47:272–76.

Taniguchi, M., and Y. Kakizawa. 2000. “Asymptotic Theory of Statistical Inference for Time Series.” New York: Springer.

Whittle, P. 1963. *Prediction and Regulation*. London: English Universities Press.

Wiener, N. 1949. *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley.