

RESEARCH REPORT SERIES
(Statistics #2012-08)

Optimal Signal Extraction with Correlated Components

Tucker McElroy
Agustin Maravall¹

¹ Bank of Spain

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: July 11, 2012

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Optimal Signal Extraction with Correlated Components

Tucker McElroy¹ and Agustin Maravall
U.S. Census Bureau and Bank of Spain

¹ Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

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 hysteresis models is presented, along with empirical results on several economic time series. Comparisons are made between signal extraction estimates from traditional WK filters and those arising from the hysteresis models.

Keywords. ARIMA, Nonstationary, Seasonality, Time Series.

Disclaimer This paper is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau.

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 – 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, Koehler, Snyder, and Gross (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), 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 an estimator that has minimum mean squared error among all estimators linear in the data, or alternatively one that has minimum mean squared error among all estimators when the data is Gaussian. So the estimators 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 hysteresis models to time series data, and generate the component estimates. The most viable approach at present is to set up a structural model for the data, 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 hysteresis structural model. Our main focus is to provide explicit signal extraction formulas and illustrate how these can be implemented for linear structural models¹. Furthermore, we develop a fundamental interpretation of the optimal hysteresis filter as a convex combination of the Beveridge-Nelson (BN) and Wiener-Kolmogorov (WK) filters². 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 hysteresis filters. A discussion of implementation of the signal extraction filters is also provided³, with a frequency domain analysis of the filters that provides further insight into the role of hysteresis. Section 4 proposes a simplistic structural hysteresis 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 hysteresis signal estimates. Section 6 summarizes our findings, and all proofs are in the Appendix.

2 Signal Extraction Formulas

We consider the additive decomposition of our data vector $Y = (Y_1, Y_2, \dots, Y_n)'$ into signal S and noise N , via $Y = S + N$. For example, the signal might be the trend component, while the noise includes the seasonal and irregular components. Following Bell (1984), we let $\{Y_t\}$ be an integrated process such that $W_t = \delta(B)Y_t$ is stationary, where 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.) Typically, it is assumed that $\{W_t\}$ is stationary, although this is not required for Theorem 1 below. 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), such that the series

$$U_t = \delta^S(B)S_t \quad V_t = \delta^N(B)N_t \quad (1)$$

are mean zero 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

¹By 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.

²Technically, 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).

³R code (R Development Core Team, 2008) for the finite sample filters is available from the first author.

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 ; since the latter operators are relatively prime, $\delta = \delta^S \cdot \delta^N$ and $d = d_S + d_N$.

Now we can write (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 $\delta^S(z)$ 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

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

where W , U , V , S , and N are column vectors of appropriate dimension, with covariance matrices Γ_W , Γ_U , etc. 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

$$\Delta = \underline{\Delta}_S \Delta_N = \underline{\Delta}_N \Delta_S. \quad (2)$$

A proof of (2) can be found in McElroy and Sutcliffe (2006). With these notations it is possible to write the Mean Square Error (MSE) linear optimal estimate \widehat{S} as a linear matrix operating on 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 $\{V_t\}$:

$$\mathbb{E}[UV'] = \Gamma_{UV}.$$

We do not assume that the multivariate process $\{(U_t, V_t)\}$ is stationary in Theorem 1 below, as no special structure of the matrix Γ_{UV} is needed. In general, Γ_{UV} will not be square. Note that

$$\Gamma_{VU} = \mathbb{E}[VU'] = \Gamma'_{UV}.$$

Now from (2) it follows that

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

Then the covariance matrices are related as follows:

$$\Gamma_W = \underline{\Delta}_S \Gamma_V \underline{\Delta}'_S + \underline{\Delta}_N \Gamma_U \underline{\Delta}'_N + \underline{\Delta}_N \Gamma_{UV} \underline{\Delta}'_S + \underline{\Delta}_S \Gamma_{VU} \underline{\Delta}'_N. \quad (3)$$

We need to assume that Γ_U , Γ_V , and Γ_W are invertible, but they need not be Toeplitz, as would be true of stationary processes. 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 The initial values $Y^* = (Y_1, Y_2, \dots, Y_d)'$ are uncorrelated with $\{U_t\}$ and $\{V_t\}$.

Theorem 1 Assume the invertibility of Γ_U , Γ_V , and Γ_W . Under Assumption A, the MSE linear optimal signal extraction filter is given by

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

where the matrices M and P are

$$\begin{aligned} M &= \Delta'_S \Gamma_U^{-1} \Delta_S + \Delta'_N \Gamma_V^{-1} \Delta_N \\ P &= \Delta'_S \Gamma_U^{-1} \Gamma_{UV} \Delta'_S - \Delta'_N \Gamma_V^{-1} \Gamma_{VU} \Delta'_N. \end{aligned}$$

The error covariance matrix Γ_ε is

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

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.

The case of a bi-infinite sample is somewhat easier to describe. Now we seek the MSE optimal estimate $\hat{S}_t = \Psi(B)Y_t$, for a signal extraction filter $\Psi(B)$. Because we seek to derive time-invariant filters, we shall now assume that $\{U_t\}$ and $\{V_t\}$ are jointly stationary. 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-spectrum $f_{UV}(\lambda) = \sum_{h \in \mathbb{Z}} \mathbb{E}[U_{t+h}V_t] e^{-ih\lambda}$. Letting $z = e^{-i\lambda}$ as an abbreviation, we have the following relations in analogy with (3):

$$f_W(\lambda) = f_U(\lambda) \delta^N(z) \delta^N(\bar{z}) + f_{UV}(\lambda) \delta^N(z) \delta^S(\bar{z}) + f_{VU}(\lambda) \delta^S(z) \delta^N(\bar{z}) + f_V(\lambda) \delta^S(z) \delta^S(\bar{z}). \quad (4)$$

Then the frequency response function (frf) for $\Psi(B)$ is given in the following result.

Theorem 2 Under Assumption A, the MSE linear optimal signal extraction filter has frf given by

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

The spectral density of the stationary error process is

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

Remark 2 Clearly $\Psi(B)$ includes noise differencing $\delta^N(B)$, but a second noise differencing $\delta^N(F)$ is not included when there is cross-correlation. This differs from the conventional case, where of course $f_{UV} = 0$ and $\Psi(B)$ contains both $\delta^N(B)$ and $\delta^N(F)$ as factors.

3 The ARIMA Case: Relation to BN and WK Filters

In this section we consider bi-infinite data drawn from ARIMA component models with the following notation borrowed from Bell and Martin (2003). The process $Y_t = S_t + N_t$ as before, with

$$Y_t = \frac{\theta(B)}{\varphi(B)} a_t \quad S_t = \frac{\theta^S(B)}{\varphi^S(B)} b_t \quad N_t = \frac{\theta^N(B)}{\varphi^N(B)} c_t. \quad (5)$$

The denominator 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 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, q^N for moving average polynomials, and p, p^S, p^N for full autoregressive polynomials (including the differencing operators). Next, we write down the various spectra of Section 2 in 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

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

which has discrete fourier transform $\rho(z)\sigma_b\sigma_c$; it follows from the Cauchy-Schwarz inequality that $|\rho(z)| \leq 1$. Then we have:

$$\begin{aligned} f_U(\lambda) &= \frac{\theta^S(z)\theta^S(\bar{z})}{\phi^S(z)\phi^S(\bar{z})}\sigma_b^2 \\ f_V(\lambda) &= \frac{\theta^N(z)\theta^N(\bar{z})}{\phi^N(z)\phi^N(\bar{z})}\sigma_c^2 \\ f_{UV}(\lambda) &= \frac{\theta^S(z)\theta^N(\bar{z})}{\phi^S(z)\phi^N(\bar{z})}\rho(\bar{z})\sigma_b\sigma_c \\ f_{VU}(\lambda) &= \frac{\theta^N(z)\theta^S(\bar{z})}{\phi^N(z)\phi^S(\bar{z})}\rho(z)\sigma_b\sigma_c. \end{aligned} \quad (6)$$

Let us suppose, for the remainder of this section, that these polynomials and spectra are fixed quantities; then one may construct various filters under potentially inconsistent assumptions (e.g., constructing the WK filter when ρ is nonzero). The optimal hysteresis (H) filter arising from application of Theorem 2 will be denoted Ψ_H , and clearly depends on the function ρ :

$$\Psi_H(z) = \frac{\theta^S(z)\theta^S(\bar{z})\varphi^N(z)\varphi^N(\bar{z})\sigma_b^2 + \theta^S(z)\theta^N(\bar{z})\varphi^N(z)\varphi^S(\bar{z})\rho(\bar{z})\sigma_b\sigma_c}{\theta(z)\theta(\bar{z})\sigma_a^2}. \quad (7)$$

This follows from substitution of the relevant spectra into Theorem 2. We observe that both the BN and WK filters (for bi-infinite data) 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)). We then obtain the following formulas⁴ for their frfs:

$$\Psi_{BN}(z) = \frac{\theta^S(z)\varphi^N(z)\sigma_b}{\theta(z)\sigma_a} = \Psi_H(z)|_{\rho \equiv 1} \quad (8)$$

$$\Psi_{WK}(z) = \frac{\theta^S(z)\theta^S(\bar{z})\varphi^N(z)\varphi^N(\bar{z})\sigma_b^2}{\theta(z)\theta(\bar{z})\sigma_a^2} = \Psi_H(z)|_{\rho \equiv 0}. \quad (9)$$

That is, holding all the polynomials fixed ahead of time, setting ρ identically to unity or zero, respectively, in the hysteresis filter will yield the BN or WK filters. A further algebraic relation among the Hysteric, BN, and WK filters is derived below. Again, we suppose that the hysteresis model holds for some correlation function ρ , but we construct the filters Ψ_{BN} and Ψ_{WK} by plugging into their formulas given above. That is, we use the polynomials θ , θ^S , and φ^N associated with the hysteresis model, but then apply them to construct the BN and WK filters *as if ρ were identically one and zero respectively*. Since Ψ_{BN} is a signal extraction filter, we may denote it also by Ψ_{BN}^S when there is the need to distinguish it from the Beveridge-Nelson noise extraction filter given by

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

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

$$\begin{aligned} \Psi_H(z) &= \Psi_{BN}^S(z) \left[\frac{\theta^S(\bar{z})\varphi^N(\bar{z})\sigma_b + \theta^N(\bar{z})\varphi^S(\bar{z})\rho(\bar{z})\sigma_c}{\theta(\bar{z})\sigma_a} \right] \\ &= \Psi_{BN}^S(z) [\Psi_{BN}^S(\bar{z}) + \rho(\bar{z})\Psi_{BN}^N(\bar{z})] \\ &= \Psi_{WK}(z) + \rho(\bar{z})\Psi_{BN}^S(z)\Psi_{BN}^N(\bar{z}). \end{aligned} \quad (10)$$

This derivation uses several facts. The first equality arises from splitting the rational function in (7) 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 $\Psi_{BN}^S(z)$.

Let us interpret this interesting relationship. When $\rho \equiv 0$, (10) 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 $\Psi_{BN}^N = 1 - \Psi_{BN}^S$, so that the right hand side of (10) produces

$$\Psi_{WK}(z) + \Psi_{BN}^S(z)(1 - \Psi_{BN}^S(\bar{z})) = \Psi_{BN}^S(z).$$

Thus the hysteresis filter is equal to the BN filter in this case, which is correct. More generally, as in the work of Proietti (2006), the innovation correlation function $\rho(\bar{z}) = \rho$ might be constant

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

in $[-1, 1]$, and then the hysteresis filter is just a weighted combination of the familiar WK and BN (signal and noise) filters. When $\rho(\bar{z})$ is non-constant, additional lag and amplitude effects are superimposed on the constituent filters.

The actual implementation of the finite sample formulas is straightforward once the spectra and cross-spectra are known. We outline this procedure below. 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.

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

Of course, step 1 requires the fitted component models. How is this to be accomplished in practice? A current approach in the literature on hysteresis uses structural models with correlated innovations, and estimates the log Gaussian likelihood using state space algorithms; see Proietti (2006). 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 hysteresis structure, with correlation derivable from other parameters of the model. In Section 4 below, we provide details on a structural hysteresis model that generalizes the work of Proietti (2006) to seasonal time series.

We next illustrate the frequency response function (frf) of hysteresis 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 $\Phi(z) = 1 - 2\kappa \cos(\omega)z + \kappa^2 z^2$. 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

⁵Standard 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.

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 frequency response functions for the hysteresis 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 Hysteric, BN, and WK filters as described above. That is, Ψ_H is computed via (7) using $\rho = \pm.5$, whereas Ψ_{BN} and Ψ_{WK} are computed using (8) and (9) respectively, essentially ignoring the true values of ρ . We have plots for $\rho = \pm.5$: see Figures 1, 2, 3, and 4.

The Squared Gain and Phase Delay are probably of chief interest. For the Hysteric Trend in Figure 1, the low-pass shape of the Squared Gain function is familiar, although in the $\rho = .5$ 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 $\rho = -.5$. The cycle filter Squared Gain functions are as expected, and there is considerable phase advance at low frequencies when $\rho = -.5$.

The Squared Gain functions for the BN Trend filter in Figure 2 are qualitatively similar to the Hysteric 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 $\rho = .5$ versus $\rho = -.5$ is the same.

Finally, we track the relations between the Hysteric, BN, and WK filters in Figure 4. This is the visual display corresponding to equation (10), where $\rho = \pm.5$. We don’t examine Squared Gain or Phase Delay comparisons, since (10) does not apply to these nonlinear functions of the frf. The Trend (real part) frfs for $\rho = -.5$ show some dramatic differences, with the Hysteric filter providing much more smoothing than the BN and WK filters. For the Cycle filters, there is an interesting negative dip in Hysteric and BN filters – associated with their asymmetry – at the low frequencies.

4 Estimation of Structural Hysteresis Models

In order to utilize the formulas of Theorem 1, it is necessary to have the autocovariance and crosscovariance functions for the components. The two most popular model-based approaches in the literature 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 data are calculated via (3). In this way Gaussian maximum likelihood estimation provides the component models, and the signal extraction formulas

can then be applied. In contrast, the decomposition approach begins with a posited model for the data process, and obtains models for the components using algebra.

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 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 hysteresis model. 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 $\delta^P(B)$ and $\delta^T(B)$ respectively, with $Z_t = \delta^P(B)P_t$ and $X_t = \delta^T(B)T_t$ the stationary differenced series. Hence we have by assumption

$$\begin{aligned}\varphi^Y(B)Y_t &= \Theta^Y(B)a_t \\ \varphi^P(B)P_t &= \Theta^P(B)b_t \\ \varphi^T(B)T_t &= \Theta^T(B)c_t\end{aligned}$$

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 auto-regressive 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 auto-regressive portions, then we use the symbol ϕ . These conventions apply to data Y , trend T , and seasonal P . In order for the structural relation to hold, we must have $\varphi^Y = \varphi^T \varphi^P$.

As usual, 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 analogue of (4):

$$\begin{aligned}\frac{\Theta^Y(z)\Theta^Y(\bar{z})}{\varphi^Y(z)\varphi^Y(\bar{z})}\sigma_a^2 &= \frac{\Theta^P(z)\Theta^P(\bar{z})}{\varphi^P(z)\varphi^P(\bar{z})}\sigma_b^2 + \frac{\Theta^T(z)\Theta^T(\bar{z})}{\varphi^T(z)\varphi^T(\bar{z})}\sigma_c^2 + \sigma_I^2 \\ &+ \frac{\Theta^P(z)}{\varphi^P(z)}\tau\sigma_b\sigma_I + \frac{\Theta^P(\bar{z})}{\varphi^P(\bar{z})}\tau\sigma_b\sigma_I + \frac{\Theta^T(z)}{\varphi^T(z)}\xi\sigma_c\sigma_I + \frac{\Theta^T(\bar{z})}{\varphi^T(\bar{z})}\xi\sigma_c\sigma_I \\ &+ \frac{\Theta^T(z)\Theta^P(\bar{z})}{\varphi^T(z)\varphi^P(\bar{z})}\rho\sigma_b\sigma_c + \frac{\Theta^P(z)\Theta^T(\bar{z})}{\varphi^P(z)\varphi^T(\bar{z})}\rho\sigma_b\sigma_c.\end{aligned}\tag{11}$$

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 (11), 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 (11) will always be non-negative.

Now (11) tells us how the structural components are related to the data 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.

Let us illustrate this through a specific class of structural models. Let $\varphi^P(z) = U(z) = 1 + z + \dots + z^{11}$, 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 (LLM) or Smooth Trend Model (STM), as well as the basic seasonal model, popularized in Harvey (1989) and utilized in the software STAMP. Here we add the novel facet of there being nonzero correlation between the three innovations.

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

$$\begin{aligned} f_W(\lambda) = & \sigma_b^2 |1 - z|^{2d} + \sigma_c^2 |U(z)|^2 + \sigma_I^2 |1 - z|^{2d} |U(z)|^2 \\ & + \tau \sigma_b \sigma_I (U(z) + U(\bar{z})) |1 - z|^{2d} + \xi \sigma_c \sigma_I \left((1 - z)^d + (1 - \bar{z})^d \right) |U(z)|^2 \\ & + \rho \sigma_b \sigma_c \left((1 - z)^d U(\bar{z}) + (1 - \bar{z})^d U(z) \right). \end{aligned} \tag{12}$$

Estimation is now straightforward in principle because the exact Gaussian likelihood for the differenced data can be easily computed from the above expression. However, the model with $d = 1$ is actually not identified, whereas the $d = 2$ case is indeed identifiable; details are provided in the Appendix. Henceforth we restrict attention to the $d = 2$ case. With the differenced data vector denoted by W with covariance matrix Σ , our scale log likelihood is

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

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 $\{W_t\}$.

We henceforth refer to this model (12) – with $d = 2$ – as the Basic Hysteric Structural Model (BHSM). The maximum likelihood estimates (MLEs) for the parameter vector ψ will be denoted by $\hat{\psi}$. Details on the computation of MLEs of the BHSM are given in the Appendix. The next section examines the BHSM on time series data.

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 89 U.S. Census Bureau monthly time series for which the Box-Jenkins airline model was selected by the automdl procedure of X-12-ARIMA. We then fit the BHSM to each of these series, looking for cases where the hysteresis model's performance rivals 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 89 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. Running the automdl option of X-12-ARIMA, accounting for transformations and fixed effects, produced a subset of 89 series⁶ 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 hysteresis 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. 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 Gaussian Likelihood Ratio (GLR) test since the models are not nested). For x42100 the AIC of the BHSM is better than that of the airline model. Both BHSMs have some full correlations, indicating redundancies among the innovation sequences⁷. 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 data evaluated at the MLEs, which will be denoted $\widehat{\Sigma}$. Then the time series residuals are

$$\epsilon = \widehat{\Sigma}^{-1/2}W,$$

⁶34 foreign trade, 4 retail, 6 housing, and 44 manufacturing.

⁷This 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.

where we can use a matrix square root or a Cholesky factor to define $\widehat{\Sigma}^{1/2}$. 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 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.

This analysis gives some evidence that the BHSM is a viable contender for the SARIMA class, and the airline model in particular, for these particular series⁸. We reiterate that we are not primarily concerned with empirical validation of the hysteresis hypothesis (see Jäger and Parkinson (1994)), but in how to estimate signals of interest when hysteresis is deemed to be present. We next compute estimates of seasonal, trend, nonseasonal, and irregular for these series, along with their standard errors. Figure 6 displays some of these results, showing the trend, nonseasonal (i.e., seasonal adjustment), and seasonal components, along with time-varying MSEs and one of the filters. The optimal signal extraction matrix F for hysteresis processes need not be centrosymmetric, 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.

6 Conclusion

This paper sets out much of the mathematical and statistical theory for hysteresis time series, taking for granted that such processes are of interest to the statistical and econometric community. Indeed, analysis of economic data utilizing hysteresis 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 finite sample contexts; (ii) linkages of hysteresis filters to more well-known BN and WK filters in the context of component ARMA 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 BSM to trend-seasonal-irregular processes having hysteresis; (v) demonstration of the BHSM performance on a Foreign Trade series of the U.S. Census Bureau and an Industrial Production series of the Federal Reserve Board.

We have not undertaken a justification of the use of hysteresis 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 hysteresis model for his data, the material in this paper will be crucially important

⁸There is an issue of data-snooping here. Starting with 89 series, we refined the list of examples until we found two series for which BHSM was superior. It is possible that 2 out of 89 series were bound to prefer BHSM to the airline model just by chance.

for understanding: (a) how to model it, (b) how to estimate it (both model parameters and the actual signal extraction estimates), and (c) how to think about the properties of hysteresis 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. Hysteric models may be appropriate for some data, in which case the tools of this paper should prove quite useful for modeling and analysis.

Acknowledgements The first author thanks Tara Sinclair for alerting him to the identifiability issue, and David Findley for comments on the paper.

Appendix

A.1 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 (5). Then the MSE optimal filters have zero error, and are given by de-correlating the data (via the filter $\varphi(B)/\theta(B)$) followed by re-correlating according to the signal's pattern, namely by the filter $\theta^S(B)/\varphi^S(B)$. Multiplying these filters yields $\theta^S(B)\varphi^N(B)/\theta(B)$. 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 estimates under the full positive correlation hypothesis, and of course the signal and noise estimates aggregate back to the original data. 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 hysteresis 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).

A.2 Proofs

Proof of Theorem 1. Following the same technique as McElroy (2008), it suffices to show that the error process $\varepsilon = \hat{S} - S$ is uncorrelated with Y . So the error process is

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

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 . Hence

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

using (3). This establishes MSE linear optimality. For the error covariance matrix, we obtain the formula by expanding $\mathbb{E}[\varepsilon \varepsilon']$ and simplifying the algebra. \square

Proof of Theorem 2. Our strategy is to demonstrate that the filter defined by $\Psi(z)$ produces a signal extraction error process that is orthogonal to the data process, implying MSE linear optimality (cf. Bell (1984)). As usual, the filter coefficients of $\Psi(z)$ are obtained by integrating against \bar{z}^j , and we represent the filter formally via $\Psi(B)$. It is clear from the given formula that $\delta^N(B)$ can be factored out, leaving $\Psi(B) = \Omega(B) \delta^N(B)$ 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 $1 - \Psi(B) = \Phi(B) \delta^S(B)$ with

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

Then $\varepsilon_t = \Psi(B) Y_t - S_t = \Omega(B) V_t - \Phi(B) U_t$. The data 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 orthogonal to 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

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

which is independent of t , but holds for all h . Thus we can take the discrete Fourier transform of this relation by summing against \bar{z}^h for $h \in \mathbb{Z}$. This translates the equation into the frequency domain, so that we can substitute z for B and \bar{z} for F , and the spectra for the autocovariances and cross-covariances. Then simple algebra, along with the above formulas for $\Omega(z)$ and $\Phi(z)$, produces $\mathbb{E}[\varepsilon_t W_{t+h}] = 0$. A second calculation produces

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

from which the stated error spectrum follows after some tedious algebra. \square

A.3 Implementation of the Hysteric Model

We give some details on the calculation of the log likelihood for the structural Hysteric model of Section 4. First, the autocovariance sequences for each of the six terms in (12) are given as follows. We denote the cross-covariance functions for differenced components as γ_{PT} , γ_{PI} , etc. (the sequences are symmetric in the index in the case of a BHSM), and we write the values at lags zero, one, two, etc.:

$$\begin{aligned} \gamma_{TT} &: 2, -1, 0, \dots \\ \gamma_{PP} &: 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0, \dots \\ \gamma_{II} &: 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1, 0, \dots \\ \gamma_{PI} &: -2, 1, 0, 0, 0, 0, 0, 0, 0, 0, -1, 1, 0, \dots \\ \gamma_{TI} &: 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1, 0, \dots \\ \gamma_{PT} &: 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, -1, 0, \dots \end{aligned}$$

when $d = 1$. Note that γ_{II} and γ_{TI} are identical, essentially because $|1 - z|^2 = (1 - z) + (1 - \bar{z})$. This means that ξ is not identified. For $d = 2$ the cross-covariances are given by

$$\begin{aligned} \gamma_{TT} &: 6, -4, 1, 0, \dots \\ \gamma_{PP} &: 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0, \dots \\ \gamma_{II} &: 4, -2, 0, 0, 0, 0, 0, 0, 0, 0, 1, -2, 1, 0, \dots \\ \gamma_{PI} &: 6, -4, 1, 0, 0, 0, 0, 0, 0, -1, 3, -3, 1, 0, \dots \\ \gamma_{TI} &: 0, -2, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, \dots \\ \gamma_{PT} &: 0, -1, 1, 0, 0, 0, 0, 0, 0, -1, 1, 0, \dots \end{aligned}$$

These row vectors only need to be multiplied by the parameters, summed, and put into a Toeplitz covariance matrix Σ . Actually, due to the simple way in which parameters enter these models, identifiability is equivalent to linear independence of the six vectors in each case. When $d = 1$ the six vectors, viewed as column vectors, form a matrix of rank 4, and hence the model is not

identifiable. But when $d = 2$ the rank is six; also when the correlations are all set to zero, the rank is three for the remaining three column vectors for $d = 1, 2$ (a known result). See Morley, Nelson, and Zivot (2003) for more discussion on identifiability.

Now, one interesting facet 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” $\vartheta = (\vartheta_1, \dots, \vartheta_6)'$ 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.

$$\begin{aligned} \ell_{11} &= \exp \vartheta_1 & \ell_{21} &= \exp \vartheta_2 & \ell_{31} &= \exp \vartheta_3 \\ \ell_{22} &= \pi \frac{\exp \vartheta_4}{1 + \exp \vartheta_4} & \ell_{32} &= \pi \frac{\exp \vartheta_5}{1 + \exp \vartheta_5} & \ell_{33} &= \pi \frac{\exp \vartheta_6}{1 + \exp \vartheta_6} \end{aligned}$$

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

$$\begin{aligned} \psi_1 &= \exp\{2\vartheta_1\} & \psi_2 &= \exp\{2\vartheta_2\} & \psi_3 &= \exp\{2\vartheta_3\} \\ \psi_4 &= \cos(\ell_{22}) & \psi_5 &= \cos(\ell_{32}) & \psi_6 &= \cos(\ell_{22}) \cos(\ell_{32}) + \sin(\ell_{22}) \sin(\ell_{32}) \cos(\ell_{33}). \end{aligned}$$

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, puts you 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 $\hat{\vartheta}$ plugged in. The matrix D follows from calculus. The first five rows make

up a diagonal matrix with entries

$$\begin{aligned}
D_{11} &= 2 \exp\{2\vartheta_1\} \\
D_{22} &= 2 \exp\{2\vartheta_2\} \\
D_{33} &= 2 \exp\{2\vartheta_3\} \\
D_{44} &= -\sin\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \frac{\pi \exp \vartheta_4}{(1 + \exp \vartheta_4)^2} \\
D_{55} &= -\sin\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \frac{\pi \exp \vartheta_5}{(1 + \exp \vartheta_5)^2}.
\end{aligned}$$

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

$$\begin{aligned}
D_{64} &= -\sin\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \frac{\pi \exp \vartheta_4}{(1 + \exp \vartheta_4)^2} \cdot \cos\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \\
&\quad + \cos\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \frac{\pi \exp \vartheta_4}{(1 + \exp \vartheta_4)^2} \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \cdot \cos\left(\frac{\pi \exp \vartheta_6}{1 + \exp \vartheta_6}\right) \\
D_{65} &= -\cos\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \frac{\pi \exp \vartheta_5}{(1 + \exp \vartheta_5)^2} \\
&\quad + \sin\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \cdot \cos\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \frac{\pi \exp \vartheta_5}{(1 + \exp \vartheta_5)^2} \cdot \cos\left(\frac{\pi \exp \vartheta_6}{1 + \exp \vartheta_6}\right) \\
D_{66} &= \sin\left(\frac{\pi \exp \vartheta_4}{1 + \exp \vartheta_4}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \cdot \sin\left(\frac{\pi \exp \vartheta_6}{1 + \exp \vartheta_6}\right) \frac{\pi \exp \vartheta_6}{(1 + \exp \vartheta_6)^2}
\end{aligned}$$

Suppose now that we wish to estimate a constrained structural hysteresis 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 GLR theory of Taniguchi and Kakizawa (2000) can be applied: we take the difference of the log likelihoods, and the asymptotic distribution is χ_r^2 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 non-zero correlations, we have

$$\begin{aligned}
\psi_4 &= \frac{\exp \vartheta_4}{1 + \exp \vartheta_4} \cos\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right) \\
\psi_5 &= \frac{\exp \vartheta_4}{1 + \exp \vartheta_4} \sin\left(\frac{\pi \exp \vartheta_5}{1 + \exp \vartheta_5}\right).
\end{aligned}$$

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.

References

- [1] Bell, W. (1984) Signal extraction for nonstationary time series. *The Annals of Statistics* **12**, 646–664.
- [2] Bell, W. (2004) On RegComponent time series models and their applications. In A.C. Harvey, S.J. Koopman, and N. Shephard (Eds.), *State Space and Unobserved Component Models: Theory and Applications*. Cambridge, UK: Cambridge University Press.
- [3] Bell, W. and Martin, D. (2003) Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. *Journal of Time Series Analysis* **25**, 603–625.
- [4] Bell, W. and Hillmer, S. (1984) Issues involved with the seasonal adjustment of economic time series. *Journal of Business and Economics Statistics* **2**, 291–320.
- [5] Beveridge, S. and Nelson, C. (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**, 151–174.
- [6] Findley, D. and Martin, D. (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.
- [7] Gersch, W. and Kitagawa, G. (1983) The prediction of time series with trends and seasonalities. *Journal of Business and Economics Statistics* **1**, 253–264.
- [8] Ghysels, E. (1987) Seasonal extraction in the presence of feedback. *Journal of Business and Economics Statistics* **5**, 191–194.
- [9] Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- [10] Harvey, A. and Jäger, A. (1993) Detrending, stylized facts and the business cycle. *Journal of Applied Econometrics* **8**, 231–247.
- [11] Harvey, A. and Trimbur, T. (2007) Trend estimation, signal-noise ratios, and the frequency of observations. In *Growth and Cycle in the Eurozone*. Basingstoke, United Kingdom: Palgrave.
- [12] Hyndman, R., Koehler, A., Snyder, R., and Grose, S. (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* **18**, 439–454.
- [13] Jäger, A. and Parkinson, M. (1994) Some evidence on hysteresis in unemployment rates. *European Economic Review* **38**, 329–342.

- [14] McElroy, T. and Sutcliffe, A. (2006) An iterated parametric approach to nonstationary signal extraction. *Computational Statistics and Data Analysis, Special Issue on Signal Extraction* **50**, 2206–2231.
- [15] McElroy, T. (2008) Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* **24**, 1-22.
- [16] Morley, J., Nelson, C., and Zivot, E. (2003) Why are Beveridge-Nelson and unobserved-component decompositions of GDP so different? *Review of Economics and Statistics* **85**, 235–243.
- [17] Oh, K., Zivot, E., and Creal, D. (2008) The relationship between the Beveridge-Nelson decomposition and other permanent-transitory decompositions that are popular in economics. *Journal of Econometrics* **146**, 207–219.
- [18] Ord, J., Koehler, A. and Snyder, R. (1997) Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association* **92**, 1621–1629.
- [19] Pinheiro, J. and Bates, D. (1996) Unconstrained parameterizations for variance-covariance matrices. *Statistics and Computing* **6**, 289–296.
- [20] Proietti, T. (2006) Trend-cycle decompositions with correlated components. *Econometric Reviews* **25**, 61–84.
- [21] R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- [22] Snyder, R. (1985) Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society, B* **47**, 272–276.
- [23] Whittle, P. (1963) *Prediction and Regulation*. London: English Universities Press.
- [24] Wiener, N. (1949) *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications*. New York: Wiley.

Airline Model	m42110	x42100
Lik	287.60	238.16
AIC	-569.20	-470.31
θ	.65 (.046)	.65 (.1198)
Θ	.56 (.053)	.65 (.1510)
σ^2	.0062 (.00052)	.012 (.0033)
BHSM	m42110	x42100
Lik	288.35	241.49
AIC	-564.69	-470.97
σ_b^2	.00028 (.00008)	.00066 (.000196)
σ_c^2	.0000066 (.0000027)	.000022 (.0000081)
σ_I^2	.0020 (.00031)	.0088 (.000887)
ρ	-.993 (.099)	-.519 (.363)
τ	.995 (.111)	-1.000 (.0000033)
ξ	-1.000 (.0149)	.519 (.362)

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

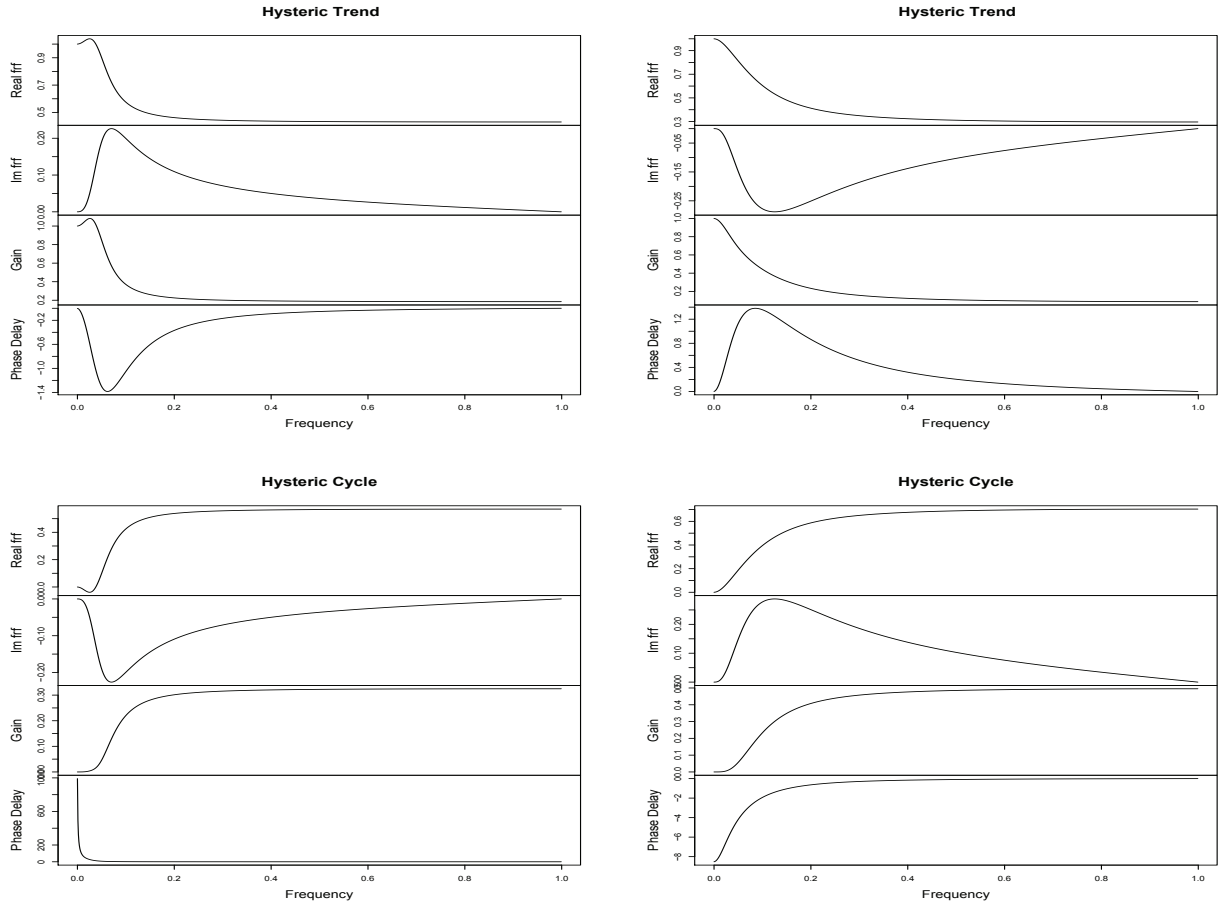


Figure 1: Frequency response functions for hysteresis trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part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

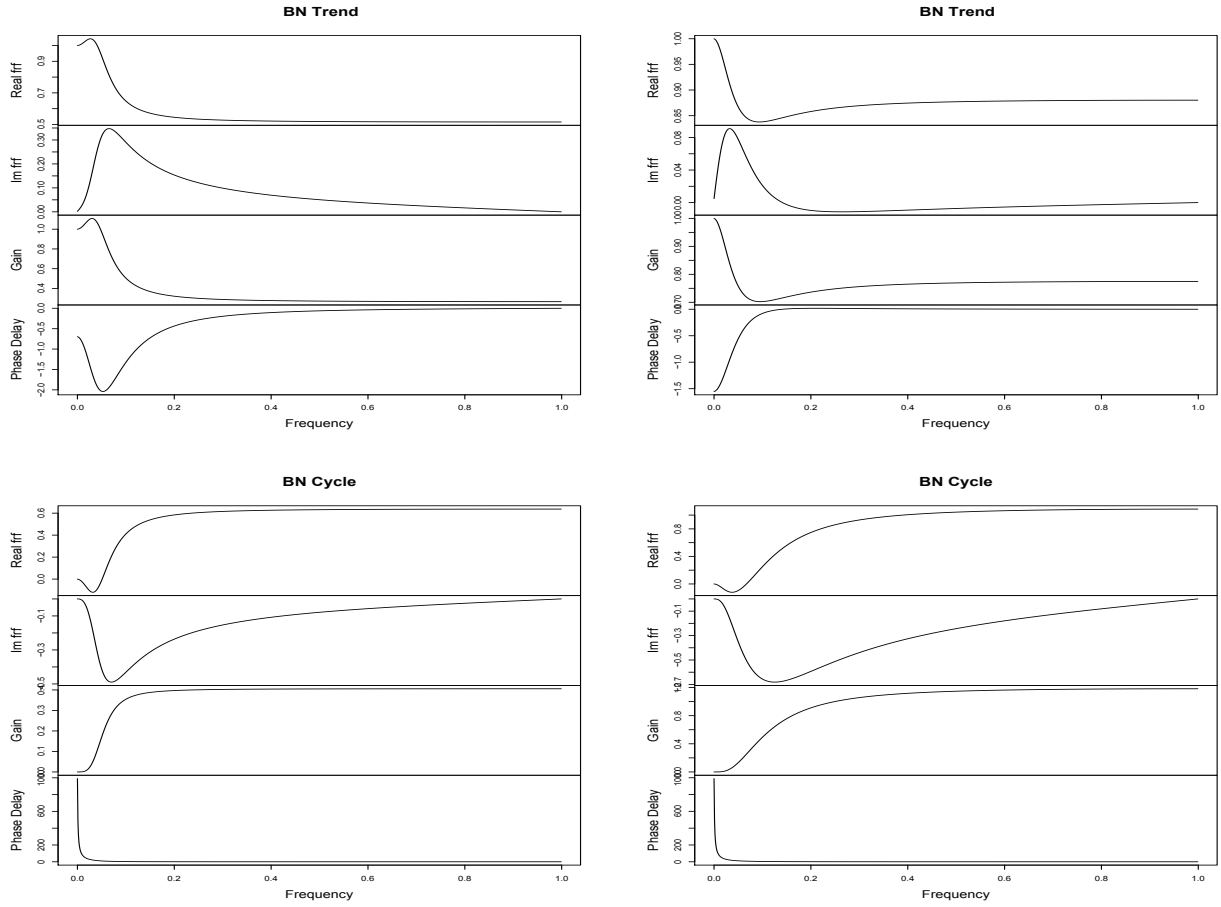


Figure 2: Frequency response functions for BN trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part frf,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

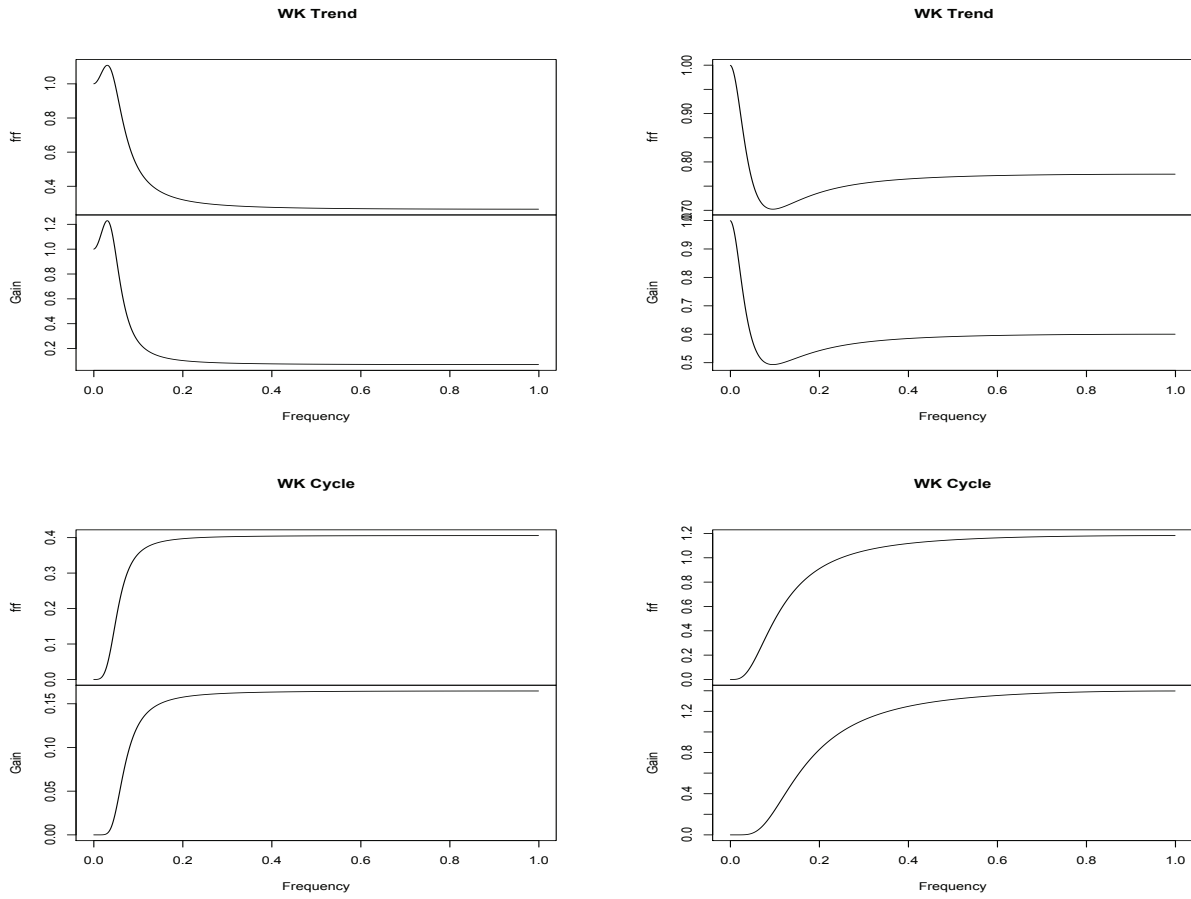


Figure 3: Frequency response functions for WK trends and cycles. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering. Within each panel, the plotted functions are real part frf, imaginary part frf,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

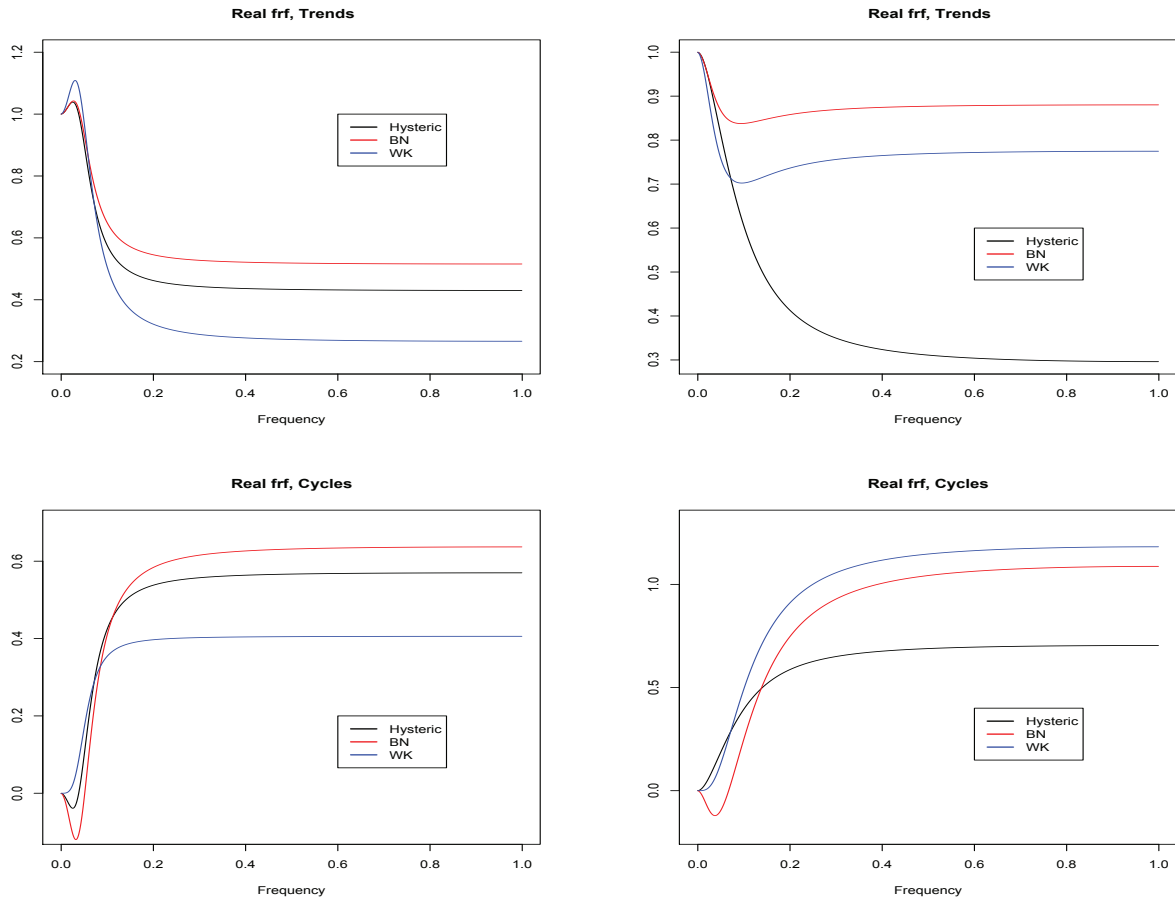


Figure 4: Real part of frequency response functions for trend and cycle filters, for hysteresis, BN, and WK approaches. Left panels are for the case $\rho = .5$, right panels for the case $\rho = -.5$. Bottom panels are for cycle filtering, top panels are for trend filtering.

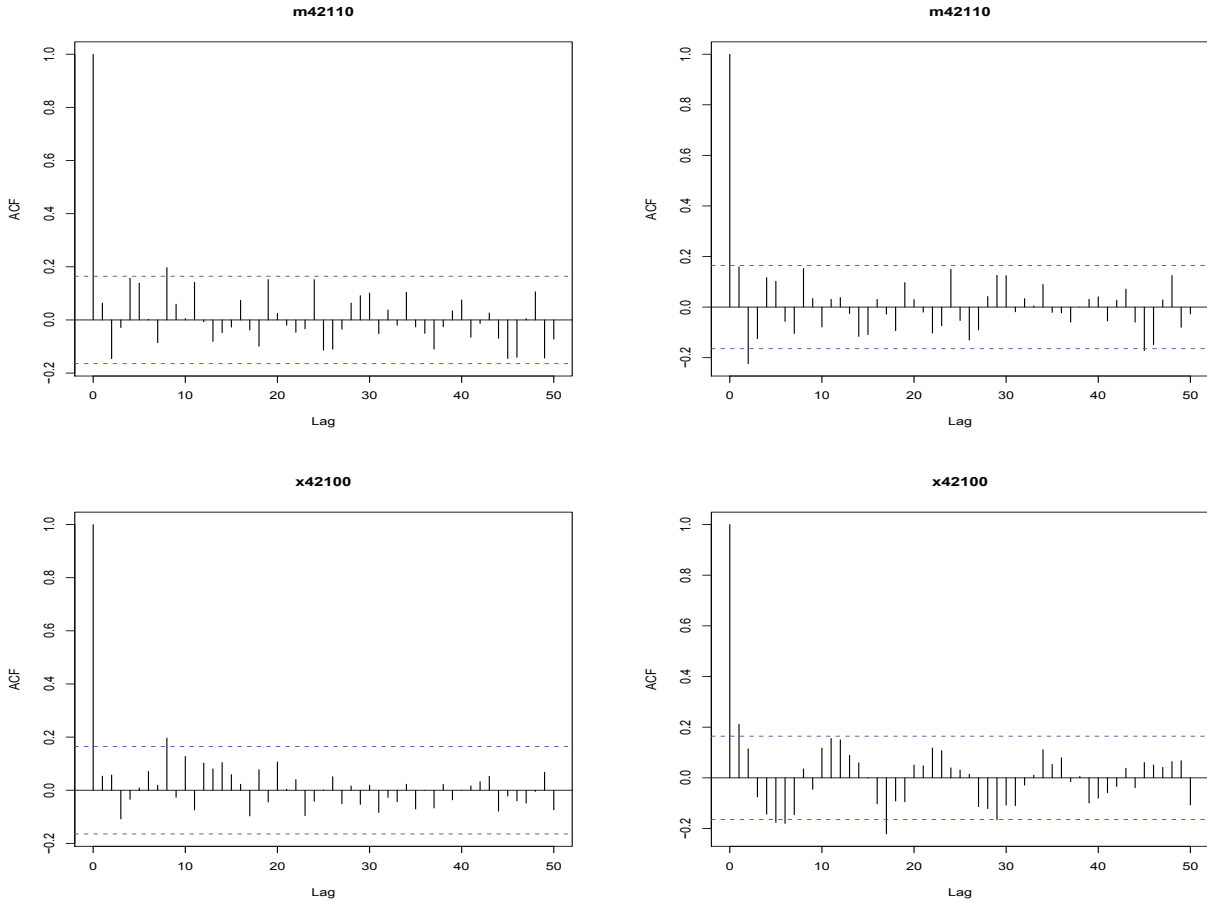


Figure 5: Sample autocorrelation plots for time series residuals. The upper panels correspond to series m42110, while the lower panels correspond to series x42100. The left panels are for the fitted airline model, whereas the right panels are for the fitted BHSM.

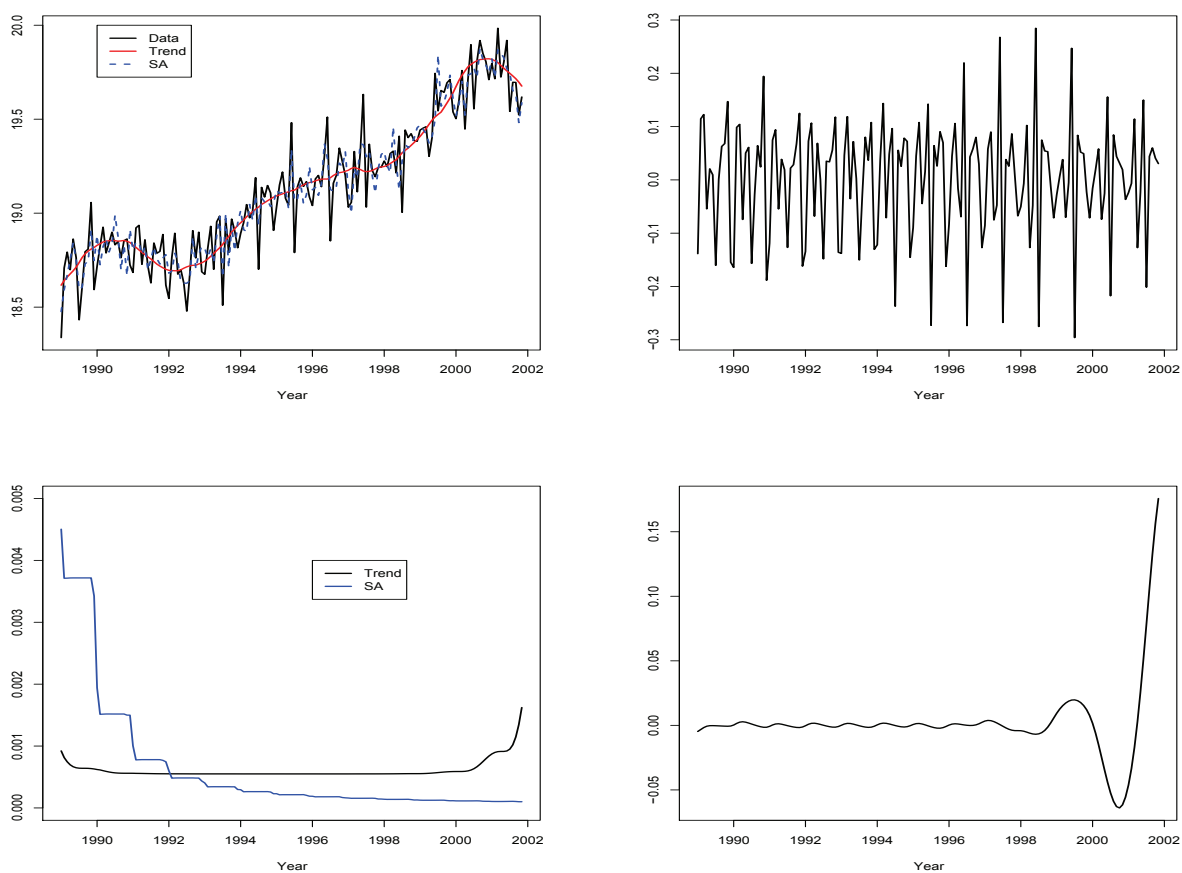


Figure 6: Signal extraction output for x42100, from the BHSM. The top left panel displays logged data along with the trend estimate and the seasonal adjustment. The bottom left panel displays time-varying MSEs for both trend and seasonal adjustment. The upper right panel displays the estimated seasonal, and the lower right panel displays the concurrent trend extraction filter weights.