

THE MULTIVARIATE LINEAR PREDICTION PROBLEM: MODEL-BASED AND DIRECT FILTERING SOLUTIONS

Abstract

Numerous contexts in macroeconomics, finance, and quality control require real-time estimation of trends, turning points, and anomalies. The real-time signal extraction problem is formulated as a multivariate linear prediction problem, the optimal solution is presented in terms of a known model, and multivariate direct filter analysis is proposed to address the more typical situation where the process' model is unknown. It is shown how general constraints – such as level and time shift constraints – can be imposed on a concurrent filter in order to guarantee that real-time estimates have requisite properties. The methodology is applied to petroleum and construction data.

Keywords: Frequency Domain, Seasonality, Time Series, Trends

1 Introduction

In the applications of time series analysis to macroeconomics, finance, and quality control it is essential to extract useful information about trends, turning points, and anomalies in real time. The practitioner does not have the luxury of sifting past data for structural breaks, indicators of regime change, or changes to volatility. Informative elections are contingent upon understanding the dynamics of various time series at time present. Because long-term movements, as well as aberrations, are defined in terms of the long-run behavior of a time series over past, present, and future, any analysis of the present state necessarily involves a degree of forecasting. This broad topic is referred to as real-time signal extraction.

A signal is any component of a time series that is deemed useful for a particular application. If long-term movements are of interest, the signal is a trend. If short-term fluctuations about a longer-term mean are of interest, the signal is a cycle. If shocks (e.g., due to rare terrorist events or natural disasters) are of interest, the signal consists of the process' extreme values. If regular patterns of an annual period, linked to cultural or meteorological patterns, are of interest, the signal is a seasonal component. If all the dynamics are of interest, but at some future time, then the signal is a multi-step ahead forecast.

However, these signals typically involves some of the past and future values of a time series – since the future is unknown, we have to rely on the present and past values. The statistical process by which a signal is estimated from available data is referred to as extraction, and the residual from the signal extraction is referred to as the noise. Whereas signals can be estimated from historical, or past, sections of a time series, when effort is focused upon time present we refer to the analysis as real-time signal extraction. Real-time signal extraction is considerably more challenging than historical signal extraction. The difficulty lies in the uncertainty about the future, which is transmitted unto the signal extraction estimates themselves.

There are other ways of viewing the real-time signal extraction problem that have been considered over the past decades: (i) constructing an asymmetric filter from a given symmetric filter, such that time delay properties are optimized; (ii) modifying a given filter to handle edge effects; (iii) constructing model-based filters such that revisions (the difference between real-time and historical estimates) are minimized. Although these approaches have yielded somewhat different bodies of literature (discussed in Section 2 below), the methodologies share common facets, since they are addressing essentially the same problem.

This article focuses upon a particular methodology called Direct Filter Analysis (DFA). As the original development of DFA (Wildi, 2008) was univariate, the methodology’s power was limited to the information content within a single time series. But because batches of time series can be closely linked, exhibiting correlated trends, common dynamics, or even predictive relationships, it is natural to expect that a multivariate extension of DFA to vector time series will more greatly facilitate informed decision making. The topic of this article is Multivariate Direct Filter Analysis (MDFA).

Many signals can be formulated as weighted linear combinations of a time series, in which case the real-time signal extraction problem can be approached as a Linear Prediction Problem (LPP). In order to pose an LPP, a solution criterion is needed, and Mean Squared Error (MSE) is often used: one seeks a real-time signal extraction that has minimal MSE discrepancy with the actual target signal. The LPP is solved by substituting multi-step ahead forecasts and backcasts for any requisite values of the process lying outside the available sample; however, this solution depends on knowing something about the dynamics in the time series process (because forecasting and backcasting formulas require the Wold decomposition to be available). The most venerable approach to understanding these dynamics is to posit a time series model, fit this model to the observed data, and plug-in the formulas for the forecasts and backcasts. This approach, which goes back to the work of Yule in the 1930s, is called the classic paradigm, being based upon a Model-Based Analysis (MBA).

An attractive feature of MBA is that analytical formulas for the LPP solutions can often be obtained, thereby facilitating computation. The philosophy underpinning the classic paradigm is that a Data Generation Process (DGP) exists, and statistical inference attempts to identify a model class for the DGP, fitting that model via estimating values of the parameters. While recognizing that any such model need not be correct, i.e., exactly match the DGP itself, such models can yet be useful to the extent to which they reflect important features in the data. Yet it is difficult to keep a model simple – which is necessary to its utility – and at the same time be sufficiently versatile to explain all the data’s features. For a given LPP, only a subset of the DGP’s features are necessary to be successful in prediction. For instance, long-term forecasting stresses the low-frequency movements of the DGP, and identification of the high-frequency dynamics is less important.

The full set of LPP solutions for a given time series is greatly constrained once a model is introduced, as only a particular subset of solutions can be obtained. If the model is badly mis-

specified, the resulting LPP solution will be inadequate. This empirical disfunctionality motivated the genesis of DFA, which essentially provides access to a much wider pool of LPP solutions. Of course, model mis-specification is always present; the issue is whether it has a significant impact upon the objectives of analysis. For instance, a given model’s mis-specification may have grave repercussions for certain problem structures, while being adequate for other LPPs. The given LPP of interest determines the gravity and impact of model mis-specification.

The overall leitmotif of this paper is to provide a mechanism for solving an LPP by adapting a given bi-infinite filter to an available finite sample. We first show this can be done by substituting forecasts and backcasts based on formulas involving the Wold decomposition of the process (our main result in Section 3 below is formulated just in terms of forecasts, as it would be applied to the case of a long sample, but Appendix A discusses the full extension to cases where backcasts are necessary). Our second goal is to demonstrate how the same solution can be obtained nonparametrically if our class of concurrent filters is sufficiently broad. Thirdly, we adapt this so-called MDFA to allow for natural constraints on the filter class.

This article presents the generalized treatment of the multivariate LPP in Section 3, following background on the overall filtering framework in Section 2. Section 4 develops MDFA in its basic form, with extensions to level and time shift constraints, and finally a general form suitable for nonstationary time series. We present applications in Section 5 to multivariate trend estimation in Petroleum data, as well as multivariate seasonal adjustment of Construction data. Section 6 summarizes our findings, with proofs in Section 7; background material and additional figures are provided in the Supplementary material.

2 Background and Framework

It is common to approach real-time problems with linear filters, although there is a burgeoning literature on non-linear techniques. Borowski, Schettlinger, and Gather (2009) examine the multivariate real-time problem using regression techniques, with extensions in Schettlinger, Fried, and Gather (2010) that robustify the results. Nonparametric approaches to signal extraction may involve singular spectrum analysis (Golyandina, Nekrutkin, and Zhigljavski (2001)), a regularized singular value decomposition (Lin, Huang, and McElroy (2019)), or wavelets, for example – see Alexandrov, Bianconcini, Dagum, Maass, and McElroy (2012) for an overview. However, in this article we are focused on real-time signal extraction problems associated with a linear filter.

There exists a substantial literature addressing the issue of how to obtain a causal (or asymmetric) adaptation of a given symmetric filter. Musgrave (1964) proposed a causal version of the Henderson trend filter (Henderson, 1916), with extensions provided in Doherty (2001), Gray and Thomson (2002), Dagum and Luati (2002), and Proietti and Luati (2008). The methodology of reproducing kernel Hilbert spaces was applied by Dagum and Bianconcini (2008) to generate a causal adaptation of the Henderson trend filter; see Dagum and Luati (2012) for an overview. Note that these are nonparametric methods that focus upon the phase-delay (i.e., time lag) properties of asymmetric filters, developed without reference to the particular dynamics of any given data set.

Baxter and King (1999) proposed an ideal band-pass filter for cycle extraction, which would suppress all frequencies in the data that do not correspond to the business cycle; a real-time version was discussed by Christiano and Fitzgerald (2003). In related work, Pollock (2009) discussed the wrapping of filter coefficients as a way to obtain asymmetric versions of a frequency-selective filter such as the ideal band-pass. Similarly, there is interest among economists in generating asymmetric versions of the Hodrick-Prescott (HP) filter described in Hodrick and Prescott (1997): see Mise, Kim, and Newbold (2005), Garratt, Lee, Mise, and Shields (2008, 2009), and Garratt, Mitchell, Vahey, and Wakerly (2011).

The HP filter can be viewed as an optimal Wiener-Kolmogorov (WK) filter for trend extraction from a process consisting of an integrated random walk plus white noise (McElroy, 2008b), and hence the optimal real-time filter (assuming this model specification is correct) would be given by the finite-sample WK filter formulas discussed in Bell and Martin (2004), Pollock (2007) and McElroy (2008a). The multivariate finite-sample WK filters are discussed in McElroy and Trimbur (2015) and McElroy (2017); also see Harvey (1989) for a state space formulation, and McElroy and McCracken (2017) for the case of multivariate forecasting. The finite-sample WK filter is a model-based approach to the real-time problem, and in contrast with the above nonparametric approaches, assumes a very specific and complete knowledge about the data's dynamics; see Findley and Martin (2006) and Tiller (2012). The discrepancy between a symmetric WK filter and its concurrent approximation is equivalent to examining the revision relationship between historical and real-time estimators (Maravall and Pérez, 2012); revisions, and the impact of model mis-specification, are discussed in McElroy and Wildi (2010).

When models are mis-specified, the finite-sample WK approach leads to sub-optimal real-time estimation. Wildi and Schips (2004) introduced the basic problem, showing that model mis-specification leads to substantial under-performance in real-time signal extraction problems.

Proietti (2005) assessed the impact of model mis-specification on the separation of trend from cycle; also see Harvey and Delle Monache (2009) and Delle Monache and Harvey (2011). Turning to forecasting, there is a much larger literature regarding the impact of model mis-specification – for example, see Schorfheide (2005), Del Negro, Schorfheide, Smets, and Wouters (2007), Aruoba, Diebold, and Scotti (2009), Clark and McCracken (2010), and Schorfheide and Song (2015). More recently, Wildi (2018) examined the real-time seasonal adjustment problem, focusing on univariate approaches.

The general problem of approximating a linear filter (whether a WK filter based on a model, or a nonparametric filter such as the ideal band-pass) with a concurrent version was formulated as an LPP in Wildi and McElroy (2016); below, we formulate the multivariate LPP and provide its solution. We begin with some notation needed to frame the problem.

Let $\{X_t\}$ be a zero mean N -dimensional weakly stationary time series, with autocovariance function (acf) defined for $h \in \mathbb{Z}$ via $\Gamma_h = \text{Cov}[X_{t+h}, X_t]$. (The jk th entry of each matrix is denoted $\Gamma_h(j, k)$.) We define the z -transform of $\{X_t\}$ via $X(z) = \sum_{j=-\infty}^{\infty} X_j z^j$. The spectral density function (sdf) is a Hermitian matrix-valued function of $\omega \in [-\pi, \pi]$, defined as the Fourier Transform (FT) of the acf, i.e., the z -transform of the acf evaluated at $z = e^{-i\omega}$:

$$F(\omega) = \sum_{h=-\infty}^{\infty} \Gamma_h e^{-i\omega h} = \Gamma(e^{-i\omega}).$$

Given a bounded sdf (i.e., each (j, k) th entry of F has bounded modulus as a function of ω), the acf can be recovered via inverse FT:

$$\Gamma_h = \langle F \rangle_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega) e^{i\omega h} d\omega, \quad (1)$$

which uses the bracket notation to define the integral of a function (of ω) multiplied by $e^{i\omega h}$, and the whole divided by 2π . Note that the integrand F is a matrix-valued function, so (1) is a compact way of expressing an element-wise integration that yields the various components of Γ_h .

We suppose that a finite-sample $\{X_1, X_2, \dots, X_T\}$ is available, from which the sample autocovariance can be computed via

$$\hat{\Gamma}_h = T^{-1} \sum_{t=1}^{T-h} X_{t+h} X_t'$$

for $h \geq 0$, and with $\hat{\Gamma}_h = \hat{\Gamma}'_{-h}$ for $h < 0$. Moreover we can compute the *Discrete Fourier Transform* (DFT) via

$$\tilde{X}(\omega) = T^{-1/2} \sum_{t=1}^T X_t e^{-i\omega t}. \quad (2)$$

This can be computed for any $\omega \in [-\pi, \pi]$, though for applications we restrict to the Fourier frequencies given by $2\pi j/T$ for any $-[T/2] \leq j \leq T - [T/2] - 1$. (This guarantees that, whether T is odd or even, there are T Fourier frequencies lying in $[-\pi, \pi]$.) The multivariate *periodogram* is defined to be the Hermitian matrix

$$\widehat{F}(\omega) = \widetilde{X}(\omega) \widetilde{X}(-\omega)'. \quad (3)$$

The periodogram furnishes a basic estimate of the spectral density F of the process. There is an empirical version of (1), where the periodogram is mapped to the sample autocovariance:

$$\widehat{\Gamma}_h = \langle \widehat{F} \rangle_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \widehat{F}(\omega) e^{i\omega h} d\omega. \quad (4)$$

The latent dynamics of $\{X_t\}$ can be revealed through the application of a multivariate linear filter Ψ ; this consists of a doubly-infinite sequence $\{\Psi_j\}$ of $N \times N$ -dimensional matrices, such that when applied to a time series $\{X_t\}$ a new time series $\{Y_t\}$ is obtained with z -transform $Y(z) = \Psi(z) X(z)$. Expanding this expression, we see that for any $t \in \mathbb{Z}$

$$Y_t = \sum_{j=-\infty}^{\infty} \Psi_j X_{t-j}. \quad (5)$$

It is convenient to have a notation for Y_t that alerts us to the fact that it has arisen through the convolution of $\{\Psi_j\}$ and $\{X_t\}$; we shall employ the short-hand $\{\Psi * X\}_t$ for Y_t in (5). The properties of a filter can be studied by setting $z = e^{-i\omega}$ in its z -transform, thereby yielding the frequency response function (frf): $\Psi(e^{-i\omega}) = \sum_{j=-\infty}^{\infty} \Psi_j e^{-i\omega j}$. Another quantity of interest is the derivative of a filter, defined by differentiating the z -transform: $\frac{d}{dz} \Psi(z) = \sum_{j=-\infty}^{\infty} j \Psi_j z^{j-1}$. This filter will be denoted by $\partial\Psi$ for short.

When $\Psi_j = 0$ for $j < 0$ the signal depends only on present and past values of the process, and hence the filter is called *causal*, or *concurrent*. In contrast, if $\Psi_j \neq 0$ for $j < 0$ then the filter depends on future values, and is not causal. A *real-time* signal \widehat{Y}_t is the output of a causal filter, as it can be computed given present data at hand, in “real time.” However, many signals of interest are defined through symmetric filters, which have the property that $\Psi_j = \Psi_{-j}$ for $j \geq 1$.

For applications, it is practical to use a causal filter, allowing for real-time signal estimates. (Such a filter requires an infinite past of data observations, or in practice a long sample; for short samples, a finite-length causal filter is needed, and the necessary adaptations of the following theory are discussed in Appendix A of the Supplement.) We seek a real-time estimate $\widehat{Y}_t = \{\widehat{\Psi} * X\}_t$, where $\widehat{\Psi}(z) = \sum_{j \geq 0} \widehat{\Psi}_j z^j$ is a causal filter that approximates Ψ on the time series of interest, i.e.,

the filter error

$$\varepsilon_t = Y_t - \hat{Y}_t = \{\Psi * X\}_t - \{\hat{\Psi} * X\}_t \quad (6)$$

should be stationary, mean zero, and have small variance. Because ε_t is an N -vector, we can design a criterion that involves minimizing the trace or determinant of its covariance matrix, for example. The quest for $\hat{\Psi}$ such that the filter error is small is called the *linear prediction problem* (LPP).

A model-based (MB) approach to the problem proceeds as follows: we can compute the optimal $\hat{\Psi}$ analytically, given knowledge of Ψ and the spectral density of $\{X_t\}$. (If the data is difference stationary, we can still solve the LPP, expressing it in terms of the differencing polynomial and the Wold decomposition of the differenced process.) These MB solutions to the LPP yield a formula for $\hat{\Psi}$, which can then be applied to generate real-time signals.

A deficiency with the MB approach is mis-specification: we must have the exact parametric form of the process' Wold decomposition. Multivariate direct filter analysis (MDFA) instead forgoes knowledge of this parametric form, and attempts to minimize $\det \text{Var}[\varepsilon_t]$ with respect to the unknown coefficients of $\hat{\Psi}$. The MDFA solution to an LPP is defined explicitly in Section 4 below, and is based on an empirical formulation of $\text{Var}[\varepsilon_t]$ involving the periodogram.

We mention two extensions of this basic MDFA. First, it may be of interest to constrain the solution $\hat{\Psi}$ in various ways – this can be done by restricting the class of causal filters. For instance, it may be of interest to ensure that $\hat{\Psi}$ and Ψ treat constants and trend lines in the same manner, leading to the level and time shift constraints. Second, the data process that we analyze may be difference-stationary, in which case the periodogram is massively biased and cannot be used as an estimator – we must modify the basic criterion, which can be accomplished by imposing generalized level and time shift constraints, as shown below.

3 Multivariate Linear Prediction Problems

We define the class of real-time estimation problems considered in this article.

Definition 1 A **target** is defined to be the output of any known linear filter acting on the data process, i.e., $\{Y_t\}$ is a target time series corresponding to a given filter Ψ acting on a given observed time series $\{X_t\}$ if and only if we can write $Y_t = \{\Psi * X\}_t$ for all integers t .

In practice, the target is specified by the analyst in accordance with their particular objectives. Below we provide some common examples.

Example 1 Multi-step Ahead Forecasting. Suppose that our goal is to forecast all of the component series h steps ahead, where $h \geq 1$ is the given *forecast lead*. Hence the target is expressed as $Y_t = X_{t+h}$ for all $t \in \mathbb{Z}$. This target corresponds to $\Psi(z) = z^{-h} I_N$, where I_N is the identity matrix of dimension N . Thus, each Ψ_ℓ is a $N \times N$ matrix, all of which are zero except Ψ_{-h} , which is given by I_N .

Example 2 Ideal Low-Pass. In order to estimate a trend from a given series, conceptually we wish to screen out all the higher frequency components in the data. With reference to the spectral representation, if $\Psi(e^{-i\omega})$ is zero for all ω in a band of the higher frequencies, then $\{Y_t\}$ will only be composed of low frequency stochastic sinusoids. The simplest way to achieve such an output is to design the frf as an indicator function (denoted with a χ), involving a steep cutoff of noise frequencies; see Baxter and King (1999). This is viewed by some as the best possible definition of trend, and hence the filter is called the ideal low-pass. Thus we have

$$\Psi(e^{-i\omega}) = \chi_{[-\mu, \mu]}(\omega) I_N$$

for some cutoff $\mu \in (0, \pi)$ that separates the pass-band from the stop-band. The coefficients are given by

$$\Psi_\ell = \frac{\sin(\ell\mu)}{\pi\ell} I_N$$

for $\ell \neq 0$ and $\Psi_0 = \mu/\pi I_N$.

Example 3 Model-Based Random Walk Trend. The Local Level Model (LLM) discussed in Harvey (1989) is capable of modeling a time series consisting of a random walk trend $\{W_t\}$ and a white noise irregular $\{Z_t\}$, such that $X_t = W_t + Z_t$. Both the multivariate trend and the irregular are driven by independent white noise processes, with respective covariance matrices Σ_W and Σ_Z , and the frf for the optimal trend extraction filter (McElroy and Trimbur, 2015) is

$$\Psi(e^{-i\omega}) = \Sigma_W [\Sigma_W + (2 - 2 \cos(\omega)) \Sigma_Z]^{-1}.$$

Example 4 Model-Based Integrated Random Walk Trend. Example 3 can be generalized to the Smooth Trend Model (STM) developed in Harvey (1989), where now the trend $\{W_t\}$ is an integrated random walk, i.e., the application of the second differencing filter $(1 - z)^2$ to $\{W_t\}$ yields a white noise process with covariance matrix Σ_W . Then the frf for the optimal trend extraction filter – which also coincides with the multivariate HP filter (cf. McElroy and Trimbur, 2015) – is given by

$$\Psi(e^{-i\omega}) = \Sigma_W \left[\Sigma_W + (2 - 2 \cos(\omega))^2 \Sigma_Z \right]^{-1}.$$

The chief difference with the frf of the LLM is that the sinusoidal factor is now squared.

Example 5 Model-Based Seasonal Adjustment. Flexible structural models were discussed in McElroy (2017), with atomic components for each distinct unit root (with any conjugate roots) in the differencing operator. For monthly data this specification is an integrated random walk trend component $\{W_t\}$ (identical to the trend discussed in Example 4), together with six atomic seasonal components that combine into a single seasonal component $\{S_t\}$ with differencing operator $U(z) = 1 + z + z^2 + \dots + z^{11}$, along with the irregular $\{Z_t\}$. Six separate covariance matrices govern the dynamics of the seasonal component, allowing for different degrees of smoothness at each of the six seasonal frequencies. The filter that suppresses the seasonal component $\{S_t\}$ and extracts trend $\{W_t\}$ and irregular $\{Z_t\}$ is a model-based seasonal adjustment filter, and is an example of a multivariate WK filter.

As we see from these examples, the targets of real-time signal extraction are features of the stochastic process that are of interest to a particular user. Targets can be *ad hoc* (cf. Example 2) or *model-based* (cf. Examples 3, 4, and 5), and may depend upon all the components of a vector process.

Definition 2 The Linear Prediction Problem (LPP) seeks a linear estimate such that the filter error (6) has mean zero, and such that the determinant of the filter error variance $\text{Var}[\varepsilon_t]$ is minimized.

The filter error variance matrix is referred to as the filter MSE. When the data process is itself causal and linear, it is possible to give an explicit solution to the LPP in terms of the Wold decomposition (Brockwell and Davis, 1991). All purely nondeterministic weakly stationary (mean zero) processes have a Wold decomposition $X_t = \sum_{\ell=0}^{\infty} \Theta_{\ell} \varepsilon_{t-\ell}$, where $\{\varepsilon_t\}$ is $\text{WN}(\Sigma)$ and $\Theta(z) = \sum_{\ell=0}^{\infty} \Theta_{\ell} z^{\ell}$. For any power series, we introduce the notation $[\Theta(z)]_a^b = \sum_{j=a}^b \Theta_j z^j$. With these preliminaries, we can state the solution to the LPP.

Proposition 1 *Suppose that $\{X_t\}$ is mean zero and weakly stationary with Wold decomposition expressed as $X_t = \sum_{\ell=0}^{\infty} \Theta_{\ell} \varepsilon_{t-\ell}$, where $\Theta(z)$ is invertible (i.e., has all zeroes outside the unit circle). Then the solution to the LPP posed by a target $Y_t = \{\Psi * X\}_t$ is given by*

$$\hat{\Psi}(z) = \sum_{\ell=0}^{\infty} \Psi_{\ell} z^{\ell} + \sum_{\ell=-1}^{-\infty} \Psi_{\ell} [\Theta(z)]_{-\ell}^{\infty} z^{\ell} \Theta(z)^{-1}. \quad (7)$$

Moreover, the MSE corresponding to this solution is given by

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \sum_{\ell, k=1}^{\infty} \Psi_{-\ell} [\Theta(e^{-i\omega})]_0^{\ell-1} \Sigma [\Theta(e^{i\omega})]_0^{k-1'} \Psi_{-k}' e^{i\omega(\ell-k)} d\omega. \quad (8)$$

Remark 1 A short intuitive proof can be given for formula (7): if we define \widehat{Y}_t such that all future values $X_{t-\ell}$ for $\ell < 0$ in the definition of Y_t are replaced by $-\ell$ -step ahead forecasts $\widehat{X}_{t-\ell}$, then the resulting filter error ε_t will be uncorrelated with $\{X_{t-\ell}, \ell \geq 0\}$. Hence, $\det \text{Var}[\varepsilon_t]$ will be minimized; the filter yielding such a \widehat{Y}_t is given by formula (7), because $\widehat{X}_{t-\ell}$ is given by $[\Theta(z)]_{-\ell}^{\infty} z^{\ell} \Theta(z)^{-1}$ acting upon $\{X_t\}$ (cf. McElroy and McCracken (2017)), so that heuristically

$$\widehat{Y}_t = \sum_{\ell=0}^{\infty} \Psi_{\ell} X_{t-\ell} + \sum_{\ell=-1}^{-\infty} \Psi_{\ell} \widehat{X}_{t-\ell} = \sum_{\ell=0}^{\infty} \Psi_{\ell} z^{\ell} X_t + \sum_{\ell=-1}^{-\infty} \Psi_{\ell} [\Theta(z)]_{-\ell}^{\infty} z^{\ell} \Theta(z)^{-1} X_t.$$

In other words, the LPP solution is given by replacing missing future observations with model-based forecasts, and then applying the target filter to the extended series – this is a multivariate extension of the approach described in Dagum (1978) for handling edge effects with a moving average filter. The full proof of the Proposition derives formula (8), which gives us a lower bound on the determinant of the MSE when we use sub-optimal proxies for $\widehat{\Psi}$.

As indicated by Remark 1, the result of Proposition 1 is chiefly useful when we know $\Theta(z)$. However, this is rarely the case in practice – it must be estimated. The classical parametric approach involves formulating a time series model, fitted using the Gaussian likelihood, and finally computing the LPP solution in terms of the fitted model. Alternatively, one might consider fitting a specified model such that the LPP MSE is minimized. A more broad nonparametric approach involves considering classes of concurrent filters and directly minimizing the LPP MSE over this class – this is the methodology of Direct Filter Analysis (DFA).

Illustration 1 VAR(1). Consider an LPP where the true process $\{X_t\}$ is a Vector Autoregression (VAR) of order 1. This process can be described via

$$X_t = \Phi X_{t-1} + \epsilon_t$$

for a matrix Φ that is stable, i.e., has all eigenvalues bounded by one in modulus (Lütkepohl, 2007). It is known that the VAR(1) has the causal representation $\Theta(z) = (1 - \Phi z)^{-1}$. Because for $\ell < 0$

$$[\Theta(z)]_{-\ell}^{\infty} = \sum_{j=-\ell}^{\infty} \Phi^j z^j = \Phi^{-\ell} z^{-\ell} (1 - \Phi z)^{-1},$$

we find that (7) reduces to

$$\widehat{\Psi}(z) = \sum_{\ell=0}^{\infty} \Psi_{\ell} z^{\ell} + \sum_{\ell=-1}^{-\infty} \Psi_{\ell} \Phi^{-\ell}. \quad (9)$$

The second term in this expression we denote by $L_{\Psi}(\Phi)$. Hence, the optimal concurrent filter is determined by applying the filter to past data and modifying the present weight Ψ_0 by adding the quantity $L_{\Psi}(\Phi)$. In the case of h -step ahead forecasting of the first time series (Example 1), $\widehat{\Psi}(z) = L_{\Psi}(\Phi) = \Phi^h$. This formula demonstrates that it is essential that Φ be stable, and if fitting a VAR(1) we must parametrize Φ such that stability is guaranteed. (Such a parametrization is discussed in Roy, McElroy, and Linton (2019).)

4 Multivariate Direct Filter Analysis

We can now discuss a more general solution to the LPP. One perspective on Proposition 1 is that it provides a particular class of concurrent filters that arise from specified models. However, so long as these models are mis-specified, the resulting concurrent filters will be sub-optimal. Therefore, it may be possible to improve performance by utilizing broader classes of concurrent filters that are not derived from a particular model. The MDFA seeks a concurrent filter $\widehat{\Psi}(z)$ that optimizes the determinant of the MSE in a given LPP.

4.1 Basic MDFA

Suppose that the causal filters of interest belong to a class \mathcal{G} described by a vector parameter ϑ belonging to a parameter manifold. Because we seek elements of \mathcal{G} that will solve an LPP, i.e., be a good concurrent approximation to Ψ , we use the notation

$$\mathcal{G} = \{\widehat{\Psi}_{\vartheta} : \vartheta \text{ belongs to a parameter space}\}. \quad (10)$$

First suppose that $\{X_t\}$ is weakly stationary with mean zero and spectral density F . The real-time estimation error is given in (6), which has mean zero and $N \times N$ variance matrix

$$E[\varepsilon_t \varepsilon_t'] = \left\langle \left[\Psi(e^{-i\omega}) - \widehat{\Psi}_{\vartheta}(e^{-i\omega}) \right] F(\omega) \left[\Psi(e^{i\omega}) - \widehat{\Psi}_{\vartheta}(e^{i\omega}) \right]' \right\rangle_0. \quad (11)$$

This suggests the criterion function $\det D_{\Psi}(\vartheta, G)$ for any Hermitian function G , defined via

$$D_{\Psi}(\vartheta, G) = \left\langle \left[\Psi(e^{-i\omega}) - \widehat{\Psi}_{\vartheta}(e^{-i\omega}) \right] G(\omega) \left[\Psi(e^{i\omega}) - \widehat{\Psi}_{\vartheta}(e^{i\omega}) \right]' \right\rangle_0. \quad (12)$$

In the following development, setting $G = F$ yields an ideal criterion based on the process, whereas setting $G = \widehat{F}$ (the periodogram) yields an empirical criterion, providing estimates that we can compute from data. Taking the determinant of (12) yields the MDFA criterion function. Given a filter class \mathcal{G} , the best possible concurrent filter is given by $\widehat{\Psi}_{\vartheta(F)}$, where $\vartheta(F)$ is a minimizer of $\det D_{\Psi}(\vartheta, F)$. This $\vartheta(F)$ is the Pseudo-True Value for the filter parameter, in analogy with the terminology for model parameters. A case of interest arises from taking a very broad class \mathcal{G} , namely let \mathcal{G} consist of all length q concurrent filters, with $\vartheta = \text{vec}[\Xi']$ and

$$\Xi = [\widehat{\Psi}_0, \widehat{\Psi}_1, \dots, \widehat{\Psi}_{q-1}]'. \quad (13)$$

So Ξ is a $qN \times N$ dimensional matrix. Then the criterion (12) can be rewritten as

$$D_{\Psi}(\vartheta, G) = \Xi' B \Xi - \Xi' A - A' \Xi + \langle \Psi(e^{-i\omega}) G(\omega) \Psi(e^{i\omega})' \rangle_0, \quad (14)$$

where

$$A' = [\langle \Psi(e^{-i\omega}) G(\omega) \rangle_0, \langle \Psi(e^{-i\omega}) G(\omega) \rangle_1, \dots, \langle \Psi(e^{-i\omega}) G(\omega) \rangle_{q-1}], \quad (15)$$

and B is a block matrix such that the jk th $N \times N$ block of B is $\langle G \rangle_{k-j}$ for $1 \leq j, k \leq q$. (Because G is Hermitian, $\langle G \rangle_{k-j}$ is real, and it follows that A is real as well.)

Proposition 2 *The minimizer of the MDFA criterion given by the determinant of (12), with respect to \mathcal{G} consisting of all length q concurrent filters, is*

$$\Xi(G) = B^{-1} A,$$

where the jk th block of B is $\langle G \rangle_{k-j}$, and A is given by (15). The minimal value is the determinant of

$$\langle \Psi(e^{-i\omega}) G(\omega) \Psi(e^{i\omega})' \rangle_0 - A' B^{-1} A. \quad (16)$$

Remark 2 To implement Proposition 2 in practice, G is given by the periodogram so that $\langle G \rangle_h = \widehat{\Gamma}_h$ by (4). It is necessary to compute A , given by (15), and we can proceed by approximating the integrals over a Riemann mesh corresponding to Fourier frequencies, i.e., for $0 \leq j \leq q-1$ compute

$$\langle \Psi(z) \widehat{F} \rangle_j \approx T^{-1} \sum_{j=-[T/2]}^{T-[T/2]-1} \Psi(e^{-i\omega_j}) \widehat{F}(\omega_j)$$

for $\omega_j = 2\pi j/T$.

Example 6 One-step Ahead Forecasting. Suppose we consider the one-step ahead forecasting of stationary time series and \mathcal{G} corresponds to all VMA filters of order q (i.e., the filter corresponds to a VMA($q - 1$) polynomial), where

$$\vartheta = \text{vec}[\widehat{\Psi}'_0, \widehat{\Psi}'_1, \dots, \widehat{\Psi}'_{q-1}].$$

With $\Psi(z) = z^{-1}$ from (12) we have

$$\begin{aligned} D_{\Psi}(\vartheta, G) &= \left\langle \left[e^{i\omega} I_N - \widehat{\Psi}_{\vartheta}(e^{-i\omega}) \right] G(\omega) \left[e^{-i\omega} I_N - \widehat{\Psi}_{\vartheta}(e^{i\omega}) \right]' \right\rangle_0 \\ &= \left\langle \left[I_N - \sum_{\ell=0}^{q-1} \widehat{\Psi}_{\ell} e^{-i\omega(\ell+1)} \right] G \left[I_N - \sum_{\ell=0}^{q-1} \widehat{\Psi}_{\ell} e^{i\omega(\ell+1)} \right]' \right\rangle_0 \\ &= \langle G \rangle_0 - 2 \Xi' \langle G \rangle_{1:q} + \Xi' B \Xi. \end{aligned}$$

Hence the optimizer is

$$\Xi(G) = B^{-1} \langle G \rangle_{1:q},$$

which is the first component of the solution to the Yule-Walker system of order q determined by G . Therefore the MDFA solution is the same as the fit of a VAR(q) using Proposition 1.

We designate the resulting prediction function $\widehat{\Psi}_{\vartheta(G)}$ as a *Linear Prediction Filter* (LPF). Again, when $G = F$ this LPF is a theoretical object, but when $G = \widehat{F}$ the LPF can be constructed directly from the sample. When \mathcal{G} is large enough to include the optimal MB filter $\widehat{\Psi}$ of Proposition 1, then $\widehat{\Psi}_{\vartheta(F)}$ corresponds to this $\widehat{\Psi}$ (assuming the model is correctly specified); in such a case, using the parametric model generally provides a more efficient estimation of $\widehat{\Psi}$, especially for small samples. However, if we construct $\widehat{\Psi}$ from parameter estimates obtained from the Whittle likelihood, then $\widehat{\Psi}_{\vartheta(\widehat{F})}$ is identical to the fitted optimal filter, and therefore is asymptotically efficient as well. Because \mathcal{G} may be too narrow to include the optimal filter (for example, the set of length q concurrent filters will not include any optimal filters with infinitely many coefficients), it may be advantageous to use a parametric model (i.e., proceed according to Proposition 1) if one is confident in having a correct specification. This paper's advocacy of MDFA is hinged upon cases where the practitioner does not have this certainty.

Illustration 2 VAR(1). Again consider a VAR(1) process, and suppose we wish to use MDFA to approximate the optimal LPP solution – even though we don't know the true dynamics. Let \mathcal{G} denote the set of moving average filters of length q , and G is the spectral density of the VAR(1);

the solution given by Proposition 2 can be compared to that of the LPP, which has the first q components given by

$$\Upsilon' = [\Psi_0 + L_\Psi(\Phi), \Psi_1, \dots, \Psi_{q-1}].$$

This is an approximate solution to the system $\Xi' B = A'$, because $\Upsilon' B$ has $j + 1$ th component, for $0 \leq j \leq q - 1$, equal to

$$\sum_{\ell=0}^{q-1} \Psi_\ell \langle G \rangle_{j-\ell} + L_\Psi(\Phi) \Gamma_j.$$

Noting that

$$L_\Psi(\Phi) \Gamma_j = \sum_{\ell=-1}^{-\infty} \Psi_{-\ell} \Phi^{-\ell} \Gamma_j = \sum_{\ell=-1}^{\infty} \Psi_{-\ell} \Gamma_{j-\ell},$$

because for a VAR(1) process $\Gamma_h = \Phi^h \Gamma_0$ when $h \geq 0$, we see that component $j + 1$ of $\Upsilon' B$ is

$$\sum_{\ell=0}^{q-1} \Psi_\ell \Gamma_{j-\ell} = [A']_{j+1} - \sum_{\ell=q}^{\infty} \Psi_\ell \Gamma_{j-\ell}.$$

As $q \rightarrow \infty$ the error term vanishes (for each j), indicating that $\Upsilon' B \approx A'$, or $\Xi \approx \Upsilon$.

To compute the quantities given in Proposition 2, and more generally to compute the MDFA criterion (12), we propose to approximate each integral by an average over Fourier frequencies. Although finer meshes could clearly be implemented, the Fourier frequency mesh is sufficient for statistical purposes – this is because when considering the asymptotic properties of linear functionals of the periodogram (i.e., weighted linear combinations of periodogram ordinates), there is no difference between averaging over Fourier frequencies or integrating over every frequency. Moreover, using the Fourier frequencies produces an empirical criterion function that is a closer approximation to the sample mean squared error, which is shown by the following heuristic arguments. Recalling that the real-time filter error $\varepsilon_t = Y_t - \hat{Y}_t$ has variance given by (11), the sample variance is

$$T^{-1} \sum_{t=1}^T \varepsilon_t \varepsilon_t' = T^{-1} \sum_{j=-[T/2]}^{T-[T/2]-1} \hat{F}_\varepsilon(\omega_j),$$

where \hat{F}_ε is the periodogram of the filter errors and $\omega_j = 2\pi j/T$ is a Fourier frequency. This equality is a discrete version of the Plancherel identity; the right hand side (with \hat{F}_X the periodogram of the process) is approximated by

$$T^{-1} \sum_{j=-[T/2]}^{T-[T/2]-1} \left[\Psi(e^{-i\omega_j}) - \hat{\Psi}(e^{-i\omega_j}) \right] \hat{F}_X(\omega_j) \left[\Psi(e^{i\omega_j}) - \hat{\Psi}(e^{i\omega_j}) \right]'$$

This is exactly the criterion $D_\Psi(\vartheta, \widehat{F}_X)$ of (12) with the integrals replaced by Riemann sums over the Fourier frequencies.

With this justification, we see that the entries of the matrix B in Proposition 2 are approximately computed via

$$B_{j,k} \approx T^{-1} \sum_{\ell=-[T/2]}^{T-[T/2]-1} G(\omega_\ell) \exp\{i(k-j)(\omega_\ell)\}$$

for $1 \leq j, k \leq T$. Moreover, for $0 \leq k \leq T-1$

$$A'_k \approx T^{-1} \sum_{\ell=-[T/2]}^{T-[T/2]-1} \Psi(e^{-i\omega_\ell}) G(\omega_\ell) e^{ik\omega_\ell},$$

where $A' = [A'_0, \dots, A'_{T-1}]$. Finally,

$$\langle \Psi(e^{-i\omega}) G(\omega) \Psi(e^{i\omega})' \rangle_0 \approx T^{-1} \sum_{\ell=-[T/2]}^{T-[T/2]-1} \Psi(e^{-i\omega_\ell}) G(\omega_\ell) \Psi(e^{i\omega_\ell})'.$$

4.2 Constrained MDFA

Various constraints upon the concurrent filter can be envisioned, and imposing such strictures results in a constrained MDFA. Writing $\Delta(z) = \Psi(z) - \widehat{\Psi}(z)$ as the discrepancy filter, we see from (6) that $E[\varepsilon_t]$ is given by the application of $\Delta(z)$ to $E[X_t]$; by Definition 2, we require that $E[\varepsilon_t] = 0$ for any LPP. If $E[X_t] = 0$ then this condition is always satisfied, but with nonzero means additional constraints on $\Delta(z)$ must be imposed, which implicitly amount to constraints on $\widehat{\Psi}(z)$. The following results are well-known (Brockwell and Davis, 1991): if $E[X_t]$ is constant but nonzero, then we require $\Delta(1) = 0$. If $E[X_t]$ is linear in t , then we require $\Delta(1) = 0$ and $\partial\Delta(1) = 0$. Hence, we obtain three fundamental types of constraints: Level Constraint (LC), Time-Shift Constraint (TSC), and Level and Time-Shift Constraint (LTSC). These are defined as follows:

$$\text{LC} : \Delta(1) = 0 \quad \text{or} \quad \Psi(1) = \widehat{\Psi}(1)$$

$$\text{TSC} : \partial\Delta(1) = 0 \quad \text{or} \quad \partial\Psi(1) = \partial\widehat{\Psi}(1)$$

$$\text{LTSC} : \Delta(1) = 0, \partial\Delta(1) = 0 \quad \text{or} \quad \Psi(1) = \widehat{\Psi}(1), \partial\Psi(1) = \partial\widehat{\Psi}(1).$$

In the case of concurrent filters of form (13), LC is accomplished by demanding that $\sum_{j=0}^{q-1} \widehat{\Psi}_j = \Psi(1)$. More generally, we consider linear constraints formulated via

$$\Xi = R\Phi + Q, \tag{17}$$

where R is $Nq \times Nr$ and Φ is $Nr \times N$ dimensional, consisting of free parameters; Q is a matrix of constants, and is $Nq \times N$ dimensional. This is not the most general formulation (we could instead work with $\text{vec}[\Xi']$), but is sufficient to describe LC, TSC, and LTSC.

Level Constraint (LC). $\sum_{j=0}^{q-1} \widehat{\Psi}_j = \Psi(1)$ implies that

$$\widehat{\Psi}_0 = \Psi(1) - \sum_{j=1}^{q-1} \widehat{\Psi}_j. \quad (18)$$

Hence $\Phi' = [\widehat{\Psi}_1, \widehat{\Psi}_2, \dots, \widehat{\Psi}_{q-1}]$ and

$$R = \begin{bmatrix} -1 & \dots & -1 \\ 1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \otimes I_N \quad Q = \begin{bmatrix} \Psi(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Time Shift Constraint (TSC). The constraint is $\partial\Psi(1) = \partial\widehat{\Psi}(1) = \sum_{j=0}^{q-1} j \widehat{\Psi}_j$, or $\widehat{\Psi}_1 = \partial\Psi(1) - \sum_{j=2}^{q-1} j \widehat{\Psi}_j$. Hence $\Phi' = [\widehat{\Psi}_0, \widehat{\Psi}_2, \dots, \widehat{\Psi}_{q-1}]$ and

$$R = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & -2 & -3 & \dots \\ 0 & 1 & 0 & \dots \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix} \otimes I_N \quad Q = \begin{bmatrix} 0 \\ \partial\Psi(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Level and Time Shift Constraint (LTSC). Take the Time Shift constraint formula for $\widehat{\Psi}_1$, and plug this into (18), to obtain

$$\begin{aligned} \widehat{\Psi}_0 &= \Psi(1) - \left(\partial\Psi(1) - \sum_{j=2}^{q-1} j \widehat{\Psi}_j \right) - \sum_{j=2}^{q-1} \widehat{\Psi}_j \\ &= \Psi(1) - \partial\Psi(1) + \sum_{j=2}^{q-1} (j-1) \widehat{\Psi}_j. \end{aligned}$$

Hence $\Phi' = [\widehat{\Psi}_2, \dots, \widehat{\Psi}_{q-1}]$ and

$$R = \begin{bmatrix} 1 & 2 & 3 & \dots \\ -2 & -3 & -4 & \dots \\ 1 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 1 \end{bmatrix} \otimes I_N \quad Q = \begin{bmatrix} \Psi(1) - \partial\Psi(1) \\ \partial\Psi(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

More generally, we can envision an LPP involving M linear constraints on each scalar filter in Ξ , taking the form $K = [J \otimes I_N] \Xi$, where J is $M \times q$ dimensional ($M < q$) and K is $NM \times N$ dimensional. (The LC, TSC, and LTSC examples all have this form.) In order to express this constraint in the form (17), we use the Q-R decomposition (Golub and Van Loan, 1996) of J , writing $J = CG\Pi$ for an orthogonal matrix C (which is $M \times M$ dimensional), a rectangular upper triangular matrix G (which is $M \times q$ dimensional), and a permutation matrix Π (which is $q \times q$ dimensional). Standard matrix software will provide the Q-R decomposition J , and should produce the rank of J as a by-product – if this rank is less than M , then there are redundancies in the constraints that should first be eliminated. Hence proceeding with a full rank J , we partition G as $G = [G_1 G_2]$ such that G_1 has M columns and G_2 has $q - M$ columns. This quantity $q - M$ corresponds to the number of free coefficient matrices, and is therefore the same as r . The Q-R decomposition guarantees that G_1 is an upper triangular matrix, and moreover it is invertible. Therefore

$$[G_1^{-1} C^{-1} \otimes I_N] K = ([I_M, G_1^{-1} G_2] \Pi \otimes I_N) \Xi,$$

and the action of Π (together with the tensor product) amounts to a block-wise permutation of the elements of Ξ . Let the output of this permutation be denoted

$$\begin{bmatrix} \bar{\Xi} \\ \underline{\Xi} \end{bmatrix} = (\Pi \otimes I_N) \Xi,$$

where $\bar{\Xi}$ is $NM \times N$ dimensional and $\underline{\Xi}$ is $Nr \times N$ dimensional. Then by substitution we can solve for $\bar{\Xi}$ in terms of $\underline{\Xi}$:

$$\bar{\Xi} = [G_1^{-1} C^{-1} \otimes I_N] K - [G_1^{-1} G_2 \otimes I_N] \underline{\Xi}.$$

Therefore we recognize the free variables $\Phi = \Xi$, and obtain R and Q in (17) via

$$R = \Pi^{-1} \begin{bmatrix} -G_1^{-1} G_2 \\ I_r \end{bmatrix} \otimes I_N$$

$$Q = \left(\Pi^{-1} \begin{bmatrix} G_1^{-1} C^{-1} \\ 0 \end{bmatrix} \otimes I_N \right) K.$$

These formulas allow one to compute the form (17) from given constraints, and an analytical solution to the resulting MDFA criterion be obtained from the following result.

Proposition 3 *The minimizer of the MDFA criterion given by the determinant of (12), with respect to \mathcal{G} consists of all length q concurrent filters subject to linear constraints of the form (17), is*

$$\Phi = [R' B R]^{-1} R' (A - B Q). \quad (19)$$

Letting $H = I_{Nq} - R [R' B R]^{-1} R' B$, the minimal value is the determinant of

$$\langle \Psi(e^{-i\omega}) G(\omega) \Psi(e^{i\omega})' \rangle_0 - A' R [R' B R]^{-1} R' A + Q' B H Q - 2 A' H Q. \quad (20)$$

For computation, we utilize the same approximations to B and A as discussed in the previous subsection, obtaining the constrained MDFA filter Ξ via (19) followed by (17).

4.3 Non-stationary MDFA

We here consider difference-stationary vector time series, which means there exists a scalar differencing polynomial $\delta(z)$ such that $\{\partial X_t\}$ given by $\{\delta * X\}_t$ is mean zero and covariance stationary. Examination of (6) indicates that the error process is not stationary unless we make certain assumptions about $\Delta(z) = \Psi(z) - \widehat{\Psi}(z)$. It is necessary that we can factor $\delta(z)$ from $\Delta(z)$, i.e., there exists $\widetilde{\Delta}(z)$ such that

$$\Delta(z) = \widetilde{\Delta}(z) \delta(z), \quad (21)$$

as otherwise we cannot guarantee that $\{\varepsilon_t\}$ will be stationary. However, (21) is sufficient to guarantee that the filter error be stationary, because

$$\varepsilon_t = \{\widetilde{\Delta} * \partial X\}_t$$

in such a case. We next discuss a set of filter constraints that guarantee (21), beginning with a lemma that discusses factorization of filters. We say a filter Ψ is absolutely convergent if $\sum_{j=-\infty}^{\infty} \|\Psi_j\| < \infty$ for a given matrix norm $\|\cdot\|$.

Proposition 4 Any linear filter Ψ can be expressed as

$$\Psi(z) = \Psi(\zeta) + (z - \zeta) \Psi^\sharp(z)$$

for any $\zeta \in \mathbb{C}$ such that $|\zeta| = 1$, and an absolutely convergent filter Ψ^\sharp , so long as $\partial\Psi$ is absolutely convergent. If in addition the filter $\partial\partial\Psi$ with z -transform $\partial\partial\Psi(z) = \sum_{j=-\infty}^{\infty} j(j-1) \Psi_j z^j$ is absolutely convergent, then there also exists an absolutely convergent filter Ψ^b such that

$$\Psi(z) = \Psi(\zeta) + \partial\Psi(\zeta) (z - \zeta) \bar{\zeta} + (z - \zeta)^2 \Psi^b(z).$$

Note that if $\Psi(\zeta) = 0$, it follows from Proposition 4 that $z - \zeta$ can be factored from $\Psi(z)$. Similarly, $(z - \zeta)^2$ can be factored from $\Psi(z)$ if $\Psi(\zeta) = \partial\Psi(\zeta) = 0$.

Definition 3 For $\omega \in [-\pi, \pi]$, a filter Ψ annihilates ω -noise of order 1 if $\Psi(e^{-i\omega}) = 0$, and annihilates ω -noise of order 2 if in addition $\partial\Psi(e^{-i\omega}) = 0$.

Hence, we have the following immediate corollary of Proposition 4.

Corollary 1 If a filter Ψ annihilates ω -noise of order 1 and $\partial\Psi$ is absolutely convergent, then

$$\Psi(z) = (z - e^{-i\omega}) \Psi^\sharp(z).$$

If a filter Ψ annihilates ω -noise of order 2, and $\partial\partial\Psi$ is absolutely convergent, then

$$\Psi(z) = (z - e^{-i\omega})^2 \Psi^b(z).$$

We can apply Corollary 1 to factor a noise-differencing polynomial $\delta^N(z)$ from $\Delta(z)$: for each ω such that the target filter Ψ annihilates ω -noise of order d , we impose the constraint that $\widehat{\Psi}$ shall have the same property, and hence $(z - e^{-i\omega})^d$ can be factored from both filters. For instance, if noise frequencies are ω_ℓ with multiplicities d_ℓ , then repeated application of Corollary 1 yields

$$\Psi(z) = \prod_{\ell} (z - e^{-i\omega_\ell})^{d_\ell} \Psi^\sharp(z) = \delta^N(z) \Psi^*(z)$$

for some residual filter Ψ^\sharp , where $\Psi^*(z) = \prod_{\ell} e^{-i\omega_\ell d_\ell} \Psi^\sharp(z)$ and $\delta^N(z) = \prod_{\ell} (1 - e^{i\omega_\ell} z)$. By imposing the same linear constraints on $\widehat{\Psi}$, we likewise obtain $\widehat{\Psi}(z) = \delta^N(z) \widehat{\Psi}^*(z)$, and hence

$$\Delta(z) = \left(\Psi^*(z) - \widehat{\Psi}^*(z) \right) \delta^N(z). \quad (22)$$

So if $\delta(z) = \delta^N(z)$, then (21) holds at once. More generally, a given process' differencing polynomial may be factored into relatively prime polynomials $\delta^N(z)$ and $\delta^S(z)$, which correspond to noise and

signal dynamics respectively – see Bell (1984) and McElroy (2008a). Many signal extraction filters Ψ have the property that they annihilate ω -noise of the appropriate order, such that $\delta^N(z)$ can be factored; in addition, the noise filter with z -transform $I_N - \Psi(z)$ has the same property with respect to the signal frequencies, i.e., $\delta^S(z)$ can be factored from $I_N - \Psi(z)$ in the same manner. Hence $I_N - \Psi(z) = \delta^S(z) \Psi^\circ(z)$ for some factor $\Psi^\circ(z)$, and imposing the same constraints on the concurrent filter yields

$$\Delta(z) = (I_N - \widehat{\Psi}(z)) - (I_N - \Psi(z)) = \left(\widehat{\Psi}^\circ(z) - \Psi^\circ(z) \right) \delta^S(z).$$

However, (22) also holds, and the roots of $\delta^S(z)$ and $\delta^N(z)$ are distinct (because the polynomials are relatively prime by assumption), and hence $\delta(z) = \delta^N(z) \delta^S(z)$ must be a factor. Therefore, $\widetilde{\Delta}(z) = (\widehat{\Psi}^\circ(z) - \Psi^\circ(z))/\delta^N(z)$, and (21) holds.

In summary, given a factorization of $\delta(z)$ into signal and noise differencing polynomials, the noise constraints and signal constraints on Ψ must also be imposed upon $\widehat{\Psi}$, and this ensures that $\{\varepsilon_t\}$ will be stationary with mean zero. If ω satisfies $\delta^N(e^{-i\omega}) = 0$, then we impose that $\widehat{\Psi}$ annihilates ω -noise of order given by the multiplicity of the root in $\delta^N(z)$. Otherwise, if ω satisfies $\delta^S(e^{-i\omega})$ then we impose that $\widehat{\Psi}(e^{-i\omega}) = \Psi(e^{-i\omega})$ (if the root is simple – if a double root, then also impose that $\partial \widehat{\Psi}(e^{-i\omega}) = \partial \Psi(e^{-i\omega})$). In practice, we must determine the real and imaginary parts of each such constraint, and write the corresponding constraints on $\widehat{\Psi}$ in the form $K = [J \otimes I_N] \Xi$ for filters of form (13), applying the methodology of the previous subsection. With these constraints in play, the formula (11) holds with $\Psi(z) - \widehat{\Psi}(z)$ replaced by $\widetilde{\Delta}(z)$ and F being the spectral density of $\{\partial X_t\}$, i.e., we define the nonstationary MDFA criterion function as $\det D_\Psi(\vartheta, G)$ for

$$\begin{aligned} D_\Psi(\vartheta, G) &= \langle \widetilde{\Delta}(e^{-i\omega}) G(\omega) \widetilde{\Delta}(e^{i\omega})' \rangle_0 \\ &= \left\langle \left[\Psi(e^{-i\omega}) - \widehat{\Psi}_\vartheta(e^{-i\omega}) \right] G(\omega) |\delta(e^{-i\omega})|^{-2} \left[\Psi(e^{i\omega}) - \widehat{\Psi}_\vartheta(e^{i\omega}) \right]' \right\rangle_0. \end{aligned} \quad (23)$$

The expression (23) utilizes (21), and employs the understanding that poles in $\delta(z)^{-1}$ are exactly canceled out by the corresponding zeros in $\Psi(z) - \widehat{\Psi}(z)$. Moreover, the ratio $(\Psi(z) - \widehat{\Psi}(z))/\delta(z) = \widetilde{\Delta}(z)$ is bounded in z , as the previous discussion guarantees. As a matter of convenience, given that the frequencies of singularity in $|\delta(e^{-i\omega})|^{-2}$ are a set of Lebesgue measure zero, calculation of $D_\Psi(\vartheta, G)$ can proceed by using the second expression, computing the numerical integration over only those frequencies where $\delta(e^{-i\omega})$ is nonzero. Whereas the theoretical filter error MSE is given by $D_\Psi(\vartheta, F)$, with F being the spectral density of $\{\partial X_t\}$, for estimation we approximate the integral over Fourier frequencies, and utilize the periodogram of the differenced data for G . Again, we omit

any contributions to the sum arising from Fourier frequencies that correspond to zeros of $\delta(e^{-i\omega})$, as such an omission only results in a loss of order T^{-1} . (The alternative is to compute the quantities $\tilde{\Delta}(e^{-i\omega})$ at Fourier frequencies, using the factorization results of Corollary 1; this is not worth the effort in practical applications.)

5 Simulations and Applications

We now apply the preceding methods to simulations and real data, exploring real-time trend extraction problems as well as seasonal adjustment, in a multivariate context. Since we wish to demonstrate that MDFA is effective for any one given time series with sufficiently long sample size, we use a single Monte Carlo in our simulation studies.

5.1 VAR(1) Specification and Trend Extraction

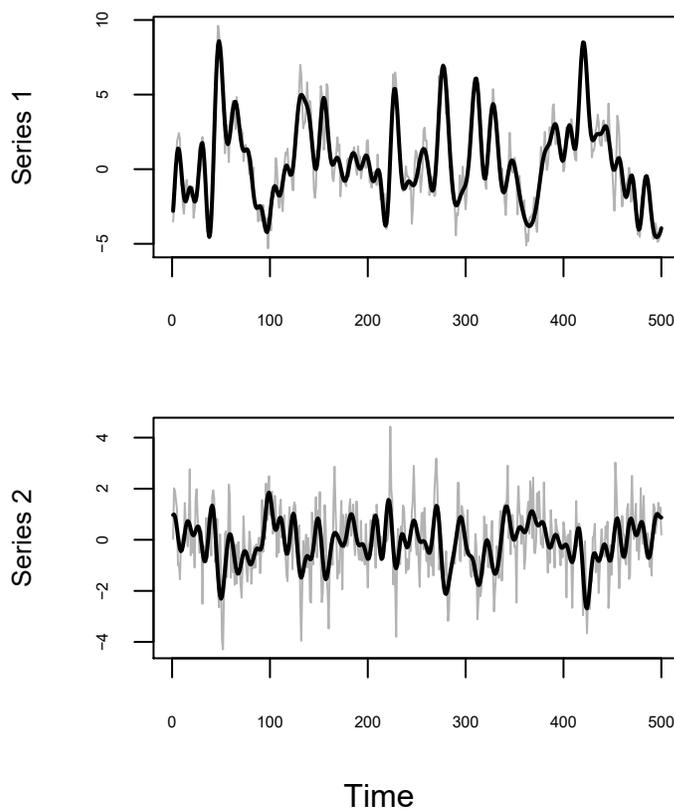


Figure 1: Bivariate low-pass trend filter applied to VAR(1) simulation (grey), with trends in black.

We suppose the true process is a VAR(1), and apply the ideal low-pass trend filter defined

in Example 2, where $\mu = \pi/6$. We seek to solve the corresponding trend extraction LPP. First, we can use the optimal solution (9) given in Illustration 1, supposing that we know that the VAR(1) parameters. Second, we can use MDFA, proceeding as if we do not know the true process is a VAR(1), as we would in practice, and hence use the periodogram; MDFA should be able to replicate the optimal solution, so long as the filter class \mathcal{G} is sufficiently rich. The VAR(1) is defined by

$$X_t = \begin{bmatrix} 1.0 & 0.5 \\ -0.2 & 0.3 \end{bmatrix} X_{t-1} + \epsilon_t,$$

with stationary initialization, and $\{\epsilon_t\}$ a Gaussian white noise of identity innovation variance. Operationally, we simulate this process with sample size 4500. Then the ideal trends are produced by truncating the MB filter to length 4001 (it is symmetric, so the indices range between -2000 and 2000) and applying to the simulation, only retaining the central 500 data points, as displayed in Figure 1. (In this way we can dispense with edge effects, and the extra 4000 observations are not used in the MDFA.) The grey lines of Figure 1 are the central 500 observations of the VAR(1) simulation, and the black line is the target. We wish to use MDFA (setting $q = 30$) with various constraints (LC, TSC, LTSC) to obtain a real-time estimate, comparing the result to the optimal solution given by implementing Illustration 1. In that case we find that

$$L_{\Psi}(\Phi) = \begin{bmatrix} 0.398 & 0.266 \\ -0.106 & 0.026 \end{bmatrix}$$

by direct calculation, and hence the optimal filter is easily computed. The in-sample MSEs of the various methods are displayed in Table 1. Note that the basic MDFA (no constraints) replicates the optimal filter, as their MSE is the same up to negligible error. When imposing a level constraint (LC) there is a modest loss to the MDFA performance, which makes sense given that the optimal filter does *not* impose a level constraint – in fact, the value of the optimal concurrent filter at frequency zero is

$$\widehat{\Psi}(1) = \begin{bmatrix} 0.982 & 0.266 \\ -0.106 & 0.610 \end{bmatrix},$$

which is quite different from I_2 . On the other hand, the time shift constraint alone (TSC) has little impact on the performance of MDFA, because $\partial \widehat{\Psi}(1) \approx 0 \cdot I_2$, i.e., the optimal filter already has this property of zero time shift. When both level and time shift constraints (LTSC) are imposed, MSE increases even more.

Series	LPP Opt	MDFA Basic	MDFA LC	MDFA TSC	MDFA LTSC
1	.3905	.3838	.3915	.3863	.4064
2	.1300	.1289	.1364	.1300	.1628

Table 1: LPP MSE for bivariate VAR(1) process – with target trend given by the LLM MB trend – for various concurrent filters: LPP Opt is the optimal filter, whereas the MDFA filters are labeled according to the constraints imposed.

5.2 Petrol

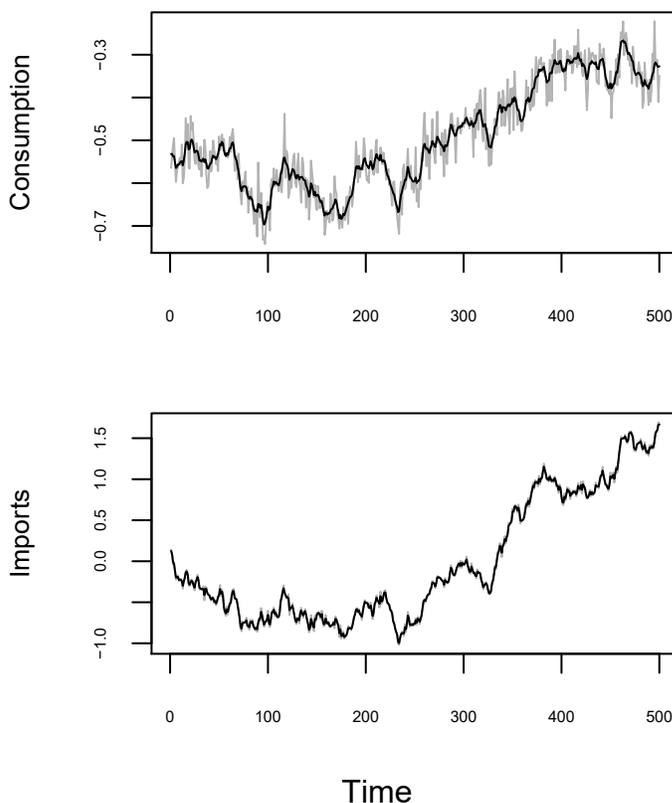


Figure 2: Bivariate trend filter applied to Null LLM simulation (grey), with trends in black. The simulation is obtained from a LLM fitted to bivariate Petrol (Consumption and Imports), so the component series of the simulation are labelled accordingly.

We now study the MB trend filter arising from a LLM (cf. Example 3) fitted to a bivariate petroleum series: Industrial Petroleum Consumption and OPEC Oil Imports, Jan 1973 through December 2016, both seasonally adjusted (displayed in Figure B.1 of the Supplement). The MLEs

for the model parameters are given by

$$\Sigma_W = \begin{bmatrix} 2.32 \cdot 10^{-4} & 5.04 \cdot 10^{-4} \\ 5.04 \cdot 10^{-4} & 34.73 \cdot 10^{-4} \end{bmatrix} \quad \Sigma_Z = \begin{bmatrix} 110.44 \cdot 10^{-5} & 7.17 \cdot 10^{-5} \\ 7.17 \cdot 10^{-5} & 128.57 \cdot 10^{-5} \end{bmatrix}.$$

Because the trend variance for the second component is 15 times larger than that of the first component, the corresponding trend filter does less smoothing.

We are interested in the performance of MDFA relative to the MB concurrent filter. We begin with a specification of the LLM exactly corresponding to the model fitted to the Petrol data, so that the MB concurrent filter $\widehat{\Psi}$ solves the LPP. (We refer to this as the null specification of the LLM.) We show that MDFA (with appropriate constraints) can replicate this optimal filter. As with the VAR(1) simulation we truncate the target MB filter to length 4001 and generate a simulated LLM of length 4500. The simulations, together with the target trends (given by the WK trend filters), are displayed in Figure 2.

	Null LLM		Alternative LLM		Petrol Data	
Series	MB	MDFA	MB	MDFA	MB	MDFA
Consumption	.1285	.1268	1.3224	.9841	.1295	.1176
Imports	.1728	.1691	.8433	.6603	2.1691	2.2033

Table 2: Empirical LPP MSE for real-time trend estimators (MB Concurrent filter versus MDFA filter) applied to bivariate LLM null process, LLM alternative process, and Petrol data, with target trend given by the null LLM MB trend. (Units of 10^{-3} .)

We apply MDFA in the manner described in Section 4.3, where $\delta^S(z) = 1 - z$ and $\delta^N(z) = 1$. By extracting the bottom row (corresponding to $T = 528$, the length of the petrol sample) of the matrix formula for the finite-sample filter (McElroy and Trimbur, 2015), we obtain a very close approximation to the MB concurrent filter, which we denote by $\widehat{\Psi}$. Results are displayed in Figure B.3 of the Appendix, with the target trend in black (these are the same black trend lines as displayed in Figure 2, but here shown without the underlying simulation) and the MDFA real-time trend in dark grey; MDFA does a good job of tracking the real trend, although with some loss to smoothness in addition to a lag effect – this is to be expected of a concurrent filter, to some degree. Indeed, the optimal concurrent filter (light grey) is exactly matched by the MDFA filter. The first column block of Table 2 shows the in-sample MSE for the two concurrent filters (MB versus MDFA), showing negligible discrepancies, i.e., the MDFA filter replicates the optimal concurrent filter.

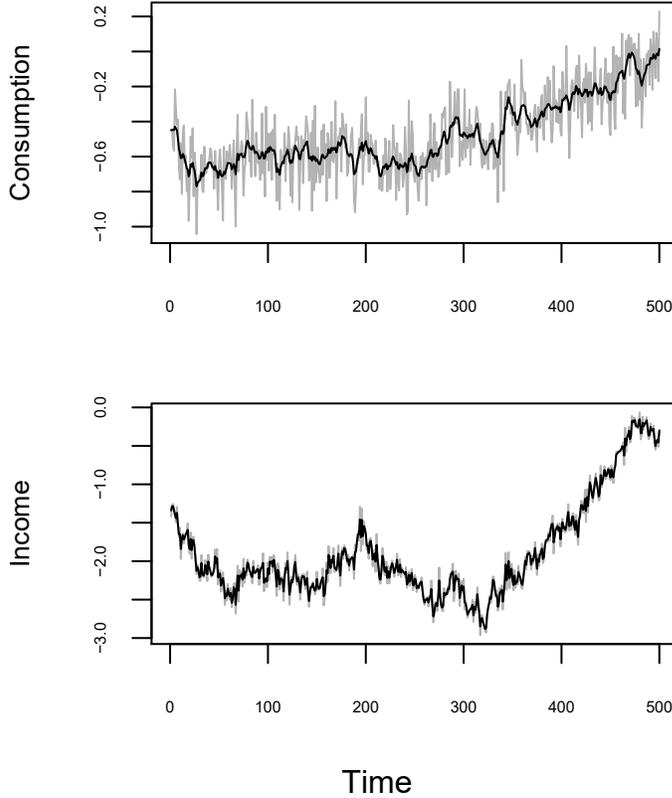


Figure 3: Bivariate trend filter applied to alternative LLM simulation (grey), with trends in black. The simulation is obtained from altering the parameters of an LLM fitted to bivariate Petrol (Consumption and Imports), so the component series of the simulation are labelled accordingly.

Next, we alter the specification to illustrate that MDFA can out-perform the MB concurrent filter. We do this by substantially increasing the variability in the irregular, producing a noisier simulation – now the MB concurrent will do too little smoothing. The new irregular covariance matrix is

$$\Sigma_Z = \begin{bmatrix} 18.32 \cdot 10^{-3} & 1.19 \cdot 10^{-3} \\ 1.19 \cdot 10^{-3} & 18.39 \cdot 10^{-3} \end{bmatrix}.$$

We refer to this as the alternative LLM process. The resulting simulation with trends is shown in Figure 3. Application of MDFA yields the trends plotted in Figure B.4 of the Appendix. For the first series, MDFA tracks the MB trend better than the MB concurrent filter does; the second block of columns in Table 2 shows the in-sample MSE for the two concurrent filters, showing substantial improvements for MDFA (26% and 22% reductions to MSE, respectively for the two series).

Finally, we conduct a comparison on the Petroleum data itself, recognizing that the LLM might be mis-specified. We do not have the long samples available considered before, so we compute a

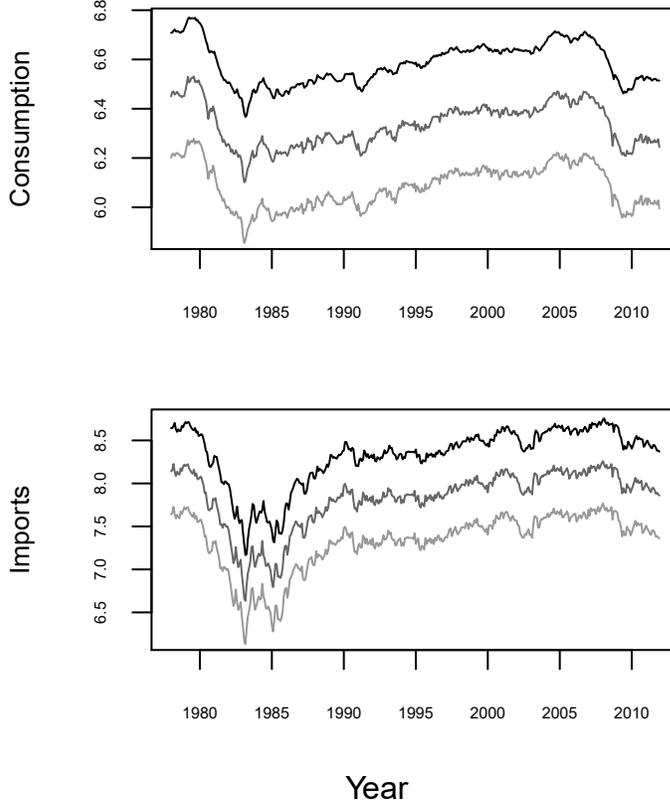


Figure 4: Monthly Petroleum data: LLM bivariate trend output (black), with non-stationary MDFA trend (dark grey) and MB concurrent trend (light grey). The trend lines have been vertically staggered for easier visualization.

series of MB trends based on the matrix formulas for signal extraction, based on the specified model, and delete the first and last five years of such trends. The result is a trend that approximately corresponds to the output of the symmetric MB filter Ψ . Taking the same MB filter as target, we compute the non-stationary MDFA solution to the LPP, and compare to the concurrent MB filter output. The resulting trends are plotted in Figure 4. The in-sample MSE is displayed in the final block of columns in Table 2, showing a modest improvement for MDFA on the first series, and similar performance on the second series.

5.3 Housing Starts

Next, we consider the quadvariate time series of Housing Starts (South, West, NorthEast, Mid-West) January 1964 through December 2012, abbreviated as Starts (displayed in Figure C.5 of the Supplement). As with the previous subsection, we simulate the null specification and verify that

MDFA can replicate the MB concurrent filter. (In order to improve finite-sample effects, we let $T = 2000$ while maintaining the filter length of 4001.) Again, we are using the constrained MDFA described in section 4.3, where the signal has differencing operator $(1 - z)^2$ (i.e., a double root at frequency zero) and the noise has differencing operator $U(z)$ (i.e., single roots at the other eleven roots of unity). The alternative specification is obtained by increasing the variability in each of the six atomic components that drive the seasonality – this has the effect of rendering the seasonal more noisy, and hence a seasonal adjustment filter should have frf with wider troughs; we therefore expect the MB concurrent filter will generate an overly stable seasonal component, leaving some dynamic seasonality behind in the seasonal adjustment. Finally, we omit the first and last fifteen years of data and engage in an empirical analysis, with results displayed in Figure C.11. All the in-sample MSEs are provided in Table 3, with results displayed in Figures C.7, C.8, C.9, and C.10 of the Appendix.

Series	Null Model		Alternative Model		Starts Data	
	MB	MDFA	MB	MDFA	MB	MDFA
South	2.1371	2.0123	9.2617	7.4844	9.7014	10.1348
West	.8051	.7398	3.5080	2.7124	2.4644	2.0859
NorthEast	.1628	.1508	.8041	.6067	1.0232	1.0411
MidWest	.4805	.4394	2.1825	1.6332	1.4035	1.1902

Table 3: Empirical LPP MSE for real-time seasonal adjustment estimators (MB Concurrent filter versus MDFA filter) applied to quadvariate structural null process, structural alternative process, and Starts data, with target seasonal adjustment given by the null structural MB seasonal adjustment.

Again, MDFA replicates (and somewhat improves upon) the MB filter in in the null specification. Under the alternative specification the MDFA filter is superior to the MB concurrent, yielding 19%, 23%, 26%, and 25% reductions to in-sample MSE, for South, West, NorthEast, and MidWest respectively. On the actual data, with fifteen years of truncation the sample is still not long enough to grant the periodogram a fully accurate portrayal of the dynamics, and therefore MDFA has worse performance for South and NorthEast, with increases of 5% and 2% to in-sample MSE, whereas for West and MidWest MDFA drops the MSE by 15% in both cases.

6 Conclusion

Real-time signal extraction – and, more generally, the LPP – is a sophisticated estimation problem with widespread applications to time series forecasting, monitoring and systems control. This paper’s approach differs from the classic time series paradigm by replicating the true structure of the multivariate LPP in the optimization criterion. Specifically, we substitute a generic target, involving a linear combination of one-step and multi-step ahead forecasts (of possibly infinite horizon) for the traditional one-step ahead error norm. Our first result is a real-time filter corresponding to a Model-Based Analysis (MBA); secondly, we generalize these solutions by adopting a nonparametric approach denoted Multivariate Direct Filter Analysis (MDFA). The generalization uses the same choice of the target signal, but utilizes a nonparametric spectral estimate in the criterion and is facilitated by a broad class of moving average filters. Various degrees of hybridization between the MBA and MDFA designs are feasible.

Our empirical examples illustrate that: (i) the MDFA is able to replicate the MBA when the chosen model corresponds to the true process, and (ii) the MDFA can outperform the MBA in cases of model mis-specification, given that the MDFA filter class is taken sufficiently large. Another distinguishing feature of our methodology is that filter coefficients are obtained directly, as an argument (i.e., a parameter) of the criterion, which allows for more precise control of frequency domain filter characteristics. As an example, improved timeliness of the real-time filter can be obtained by imposing a vanishing time-shift at frequency zero (described as the TSC of Section 4.2). More generally, we discuss the construction of filter constraints that can accommodate unit roots of arbitrary argument (i.e., complex roots of unit modulus and possibly non-zero phase) and order in the data generating process. The empirical examples demonstrate the flexibility of this approach, allowing us to address a nuanced presentation of seasonality, as well as trend extraction of varying degrees of smoothness.

Our treatment readily extends to co-integration: McElroy (2017) discusses how co-integration at a unit root $\zeta = e^{i\omega}$ can be viewed as the occurrence of rank reduction in the spectral density F of the differenced process at frequency ω . Curiously, it can occur that signal extraction filters no longer have the property that they equal zero at a noise frequency ω , if this corresponds to a co-integrating frequency. Hence, in order to construct real-time filters with the appropriate properties, instead of insisting that $\widehat{\Psi}(e^{-i\omega}) = 0$ we can just impose a mimicry of the target filter, i.e., $\widehat{\Psi}(e^{-i\omega}) = \Psi(e^{-i\omega})$. This looks exactly the same as the constraint at the signal frequency, and

is therefore trivial to implement. While in principle this extension is simple, we have omitted an application from this paper, as a full treatment of the topic seems to merit a separate article.

Another extension is motivated from the univariate DFA, which was extended in Wildi and McElroy (2019) to a still more general error criterion allowing for customization of filters, so that a practitioner can directly accomodate specific user-priorities of having a smoother real-time estimate of the signal, versus having a more timely estimate (i.e., less phase delay). A corresponding multivariate extension of the so-called Accuracy-Timeliness-Smoothness trilemma is currently under investigation by the authors.

7 Proofs

Proof of Proposition 1. In order for a linear solution to be MSE optimal, it is sufficient that the resulting error process be uncorrelated with the present and past data, denoted X_t . If we can show that the real-time signal extraction error process $\{\varepsilon_t\}$ depends only on future innovations, then by the causality of $\{X_t\}$ the error process must be uncorrelated with X_t , establishing optimality. The filter error of the putative solution is given by

$$\begin{aligned}\Psi(z) - \widehat{\Psi}(z) &= \sum_{\ell=-1}^{-\infty} \Psi_{\ell} z^{\ell} \left(1 - [\Theta(z)]_{-\ell}^{\infty} \Theta(z)^{-1}\right) \\ &= \sum_{\ell=-1}^{-\infty} \Psi_{\ell} z^{\ell} [\Theta(z)]_0^{-(\ell+1)} \Theta(z)^{-1}.\end{aligned}$$

Applying this to $\{X_t\}$ yields

$$\varepsilon_t = \sum_{\ell=1}^{\infty} \Psi_{-\ell} [\Theta(z)]_0^{\ell-1} \epsilon_{t+\ell}.$$

Noting that $[\Theta(z)]_0^{\ell-1}$ is an order $\ell-1$ polynomial in z , and is applied to $\epsilon_{t+\ell}$, it is apparent that ε_t is a linear function of future innovations $\{\epsilon_{t+1}, \epsilon_{t+2}, \dots\}$. Computing the variance of ε_t yields the expression for the minimal MSE. \square

Proof of Proposition 2. First note that the typical component of A has the form

$$\langle \Psi(z) G \rangle_{\ell} = \sum_{k=-\infty}^{\infty} \Psi_k \langle G \rangle_{\ell-k} \tag{24}$$

for $0 \leq \ell < q$, which shows that A is real-valued. The argument follows the same method as in McElroy and Findley (2015); each entry of the matrix objective function is a quadratic in Ξ , and therefore the minimizer is obtained by computing the gradient and Hessian, which are $-2A + 2B\Xi$ and $2B$ respectively, yielding the solution. Plugging back into D_{Ψ} yields (16). \square

Proof of Proposition 3. Substituting (17) in (14) yields

$$\begin{aligned} D_{\Psi}(\vartheta, G) &= \Phi' [R' B R] \Phi + [Q' B R - A' R] \Phi + \Phi' [R' B Q - R' A] \\ &\quad + Q' B Q - Q' A - A' Q + \langle \Psi(e^{-i\omega}) G(\omega) \Psi(e^{i\omega})' \rangle_0. \end{aligned}$$

Now by applying the method of proof in Proposition 2, we obtain the formula (19) for Φ . Plugging back into $D_{\Psi}(\vartheta, G)$ yields the minimal value (20). \square

Proof of Proposition 4. We claim that $\Psi^{\sharp}(z) = \sum_{j=-\infty}^{\infty} \Psi_j^{\sharp} z^j$ with

$$\Psi_j^{\sharp} = \begin{cases} \zeta^{-(j+1)} \sum_{k=j+1}^{\infty} \Psi_k \zeta^k & j \geq 0 \\ -\zeta^{-(j+1)} \sum_{k=-j}^{\infty} \Psi_{-k} \zeta^{-k} & j \leq -1. \end{cases}$$

To show this, first observe that

$$\Psi(z) - \Psi(\zeta) = \sum_{j=1}^{\infty} \Psi_j (z^j - \zeta^j) + \sum_{j=-1}^{-\infty} \Psi_j (z^j - \zeta^j).$$

Beginning with the first term, so that $j \geq 1$, we write $z^j - \zeta^j = \zeta^j (z/\zeta - 1) p_{j-1}(z/\zeta)$ where $p_k(z) = \sum_{\ell=0}^k z^{\ell}$. Next, by coefficient matching we can verify that

$$\begin{aligned} \sum_{j=1}^{\infty} \Psi_j (z^j - \zeta^j) &= (z/\zeta - 1) \sum_{j=1}^{\infty} \Psi_j \zeta^j p_{j-1}(z/\zeta) \\ &= (z/\zeta - 1) \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \Psi_k \zeta^k (z/\zeta)^j = (z - \zeta) \sum_{j=0}^{\infty} \Psi_j^{\sharp} z^j. \end{aligned}$$

Next, take $j \leq -1$, and use the symbol $y = z^{-1}$:

$$\begin{aligned} \sum_{j=-1}^{-\infty} \Psi_j (z^j - \zeta^j) &= \sum_{j=1}^{\infty} \Psi_{-j} (y^j - \zeta^{-j}) = (y\zeta - 1) \sum_{j=1}^{\infty} \Psi_{-j} \zeta^{-j} p_{j-1}(y\zeta) \\ &= (y\zeta - 1) \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} \Psi_{-k} \zeta^{-k} (y\zeta)^j \\ &= -y(z - \zeta) \sum_{j=1}^{\infty} \zeta^{j-1} \sum_{k=j}^{\infty} \Psi_{-k} \zeta^{-k} y^{j-1} \\ &= (z - \zeta) \sum_{j=-1}^{-\infty} \Psi_j^{\sharp} z^j. \end{aligned}$$

This establishes algebraically that $\Psi^{\sharp}(z)$ with coefficients as defined above equals $(\Psi(z) - \Psi(\zeta))/(z - \zeta)$, whenever the Laurent series converges. Based on the above calculations, we can write

$$\frac{\Psi(z) - \Psi(\zeta)}{z - \zeta} = \sum_{j=1}^{\infty} (\Psi_j \zeta^{j-1} p_{j-1}(z/\zeta) - \Psi_{-j} \zeta^{-j} y p_{j-1}(y\zeta)).$$

To check the absolute convergence, it suffices to set $z = 1$; note that $|p_k(\zeta)| \leq (k + 1)$ if $|\zeta| = 1$. Thus we obtain the bound

$$\left\| \frac{\Psi(z) - \Psi(\zeta)}{z - \zeta} \right\| \leq \sum_{j=1}^{\infty} j (\|\Psi_j\| + \|\Psi_{-j}\|),$$

which is finite by the assumption that $\partial\Psi$ is absolutely convergent. Next, we claim that $\Psi^b(z) = \sum_{j=-\infty}^{\infty} \Psi_j^b z^j$ with

$$\Psi_j^b = \begin{cases} \zeta^{-(j+2)} \sum_{k=j+2}^{\infty} (k-1-j) \Psi_k \zeta^k & j \geq 0 \\ \zeta^{-(j+2)} \sum_{k=-j}^{\infty} (k+j+1) \Psi_{-k} \zeta^{-k} & j \leq -1. \end{cases}$$

To verify this, observe that

$$\begin{aligned} \Psi(z) - \Psi(\zeta) - \partial\Psi(\zeta) (z - \zeta) \zeta^{-1} &= \sum_{j=1}^{\infty} \Psi_j [(z^j - \zeta^j) - j \zeta^{j-1} (z - \zeta)] \\ &\quad + \sum_{j=-1}^{-\infty} \Psi_j [(z^j - \zeta^j) - j \zeta^{j-1} (z - \zeta)]. \end{aligned}$$

First assuming that $j \geq 1$, note that $p_{\ell-1}(z) - \ell$ equals zero unless $\ell \geq 2$, and otherwise equals $\sum_{k=1}^{\ell-1} p_{k-1}(z) (z-1)$. Therefore

$$\begin{aligned} \sum_{j=1}^{\infty} \Psi_j [(z^j - \zeta^j) - j \zeta^{j-1} (z - \zeta)] &= (z - \zeta) \sum_{j=1}^{\infty} \Psi_j \zeta^{j-1} [p_{j-1}(z/\zeta) - j] \\ &= (z - \zeta)^2 \sum_{j=2}^{\infty} \Psi_j \zeta^{j-2} \sum_{k=1}^{j-1} p_{k-1}(z/\zeta) \\ &= (z - \zeta)^2 \sum_{j=0}^{\infty} \Psi_j^b z^j \end{aligned}$$

by coefficient matching in the final step. Similarly, letting $j \leq -1$ and using $z p_{\ell-1}(z) - \ell = (z-1) \sum_{k=1}^{\ell} p_{k-1}(z)$, we have

$$\begin{aligned} \sum_{j=-1}^{-\infty} \Psi_j [(z^j - \zeta^j) - j \zeta^{j-1} (z - \zeta)] &= (z - \zeta) \sum_{j=1}^{\infty} \Psi_{-j} \zeta^{-j} [-y p_{j-1}(y\zeta) + j \zeta^{-1}] \\ &= (z - \zeta)^2 y \sum_{j=1}^{\infty} \Psi_{-j} \zeta^{-(j+1)} \sum_{k=1}^j p_{k-1}(y\zeta) \\ &= (z - \zeta)^2 y \zeta^{-1} \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} (k-j) \Psi_{-k} \zeta^{-k} (y\zeta)^j \\ &= (z - \zeta)^2 \sum_{j=-1}^{-\infty} \Psi_j^b z^j \end{aligned}$$

by matching coefficients. To establish convergence of the Laurent series for Ψ^b , observe that

$$\frac{\Psi(z) - \Psi(\zeta) - \partial\Psi(\zeta)(z - \zeta)\zeta^{-1}}{(z - \zeta)^2} = \sum_{j=2}^{\infty} \Psi_j \zeta^{j-2} \sum_{k=1}^{j-1} p_{k-1}(z/\zeta) + y \sum_{j=1}^{\infty} \Psi_{-j} \zeta^{-(j+1)} \sum_{k=1}^j p_{k-1}(y\zeta).$$

Hence the matrix norm has the bound (setting $z = 1$ and taking $|\zeta| = 1$) of

$$\left\| \frac{\Psi(z) - \Psi(\zeta) - \partial\Psi(\zeta)(z - \zeta)\zeta^{-1}}{(z - \zeta)^2} \right\| \leq \sum_{j=2}^{\infty} \|\Psi_j\| \binom{j}{2} + \sum_{j=1}^{\infty} \|\Psi_{-j}\| \binom{j+1}{2},$$

using $|\sum_{k=1}^j p_{k-1}(\zeta)| \leq \binom{j+1}{2}$. Because $\partial\partial\Psi$ is absolutely convergent, the above norm is finite.

□

References

- [1] Alexandrov, T., Bianconcini, S., Dagum, E., Maass, P., and McElroy, T. (2012) The review of some modern approaches to the problem of trend extraction. *Econometric Reviews* **31**, 593–624.
- [2] Aruoba, S.B., Diebold, F.X. and Scotti, C. (2009) Real-time measurement of business conditions. *Journal of Business & Economic Statistics* **27**(4), 417–427.
- [3] Baxter, M. and King, R. (1999) Measuring business cycles: approximate bandpass filters for economic time series. *Review of Economics and Statistics* **81**, 575–593.
- [4] Bell, W. (1984) Signal extraction for nonstationary time series. *The Annals of Statistics* **12**, 646–664.
- [5] Bell, W. and Martin, D. (2004) Computation of asymmetric signal extraction filters and mean squared error for ARIMA component models. *Journal of Time Series Analysis* **25**, 603–625.
- [6] Borowski, M., Schettlinger, K. and Gather, U. (2009) Multivariate real-time signal extraction by a robust adaptive regression filter. *Communications in Statistics – Simulation and Computation* **38**(2), 426–440.
- [7] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods*. New York: Springer.
- [8] Christiano, L.J. and Fitzgerald, T.J. (2003) The band pass filter. *International economic review* **44**(2), 435–465.
- [9] Clark, T.E. and McCracken, M.W. (2010) Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics* **25**(1), 5–29.

- [10] Dagum, E.B. (1978) Modelling, forecasting and seasonally adjusting economic time series with the X-11 ARIMA method. *Journal of the Royal Statistical Society, Series D (The Statistician)* **27**(3/4), 203–216.
- [11] Dagum, E.B. and Bianconcini, S. (2008) The Henderson smoother in reproducing kernel Hilbert space. *Journal of Business & Economic Statistics* **26**(4), 536–545.
- [12] Dagum, E.B. and Luati, A. (2002) Global and local statistical properties of fixed-length non-parametric smoothers. *Statistical Methods and Applications* **11**(3), 313–333.
- [13] Dagum, E. and Luati, A. (2012) Asymmetric filters for trend-cycle estimation. In *Economic Time Series: Modeling and Seasonality*, eds. Bell, W., Holan, S., McElroy, T. CRC Press. Boca Raton, FL.
- [14] Del Negro, M., Schorfheide, F., Smets, F. and Wouters, R. (2007) On the fit of new Keynesian models. *Journal of Business & Economic Statistics* **25**(2), 123–143.
- [15] Delle Monache, D. and Harvey, A.C. (2011) The effect of misspecification in models for extracting trends and cycles. Working Papers EWP 2011/013, Euroindicators.
- [16] Doherty, M. (2001) Applications: The Surrogate Henderson Filters in X-11. *Australian & New Zealand Journal of Statistics* **43**(4), 385–392.
- [17] Findley, D.F. and Martin, D.E. (2006) Frequency domain analyses of SEATS and X-11/12-ARIMA seasonal adjustment filters for short and moderate-length time series. *Journal of Official Statistics* **22**(1), 1–34.
- [18] Garratt, A., Lee, K., Mise, E. and Shields, K. (2008) Real-time representations of the output gap. *The Review of Economics and Statistics* **90**(4), 792–804.
- [19] Garratt, A., Lee, K., Mise, E. and Shields, K. (2009) Real time representation of the UK output gap in the presence of model uncertainty. *International Journal of Forecasting* **25**(1), 81–102.
- [20] Garratt, A., Mitchell, J., Vahey, S.P. and Wakerly, E.C. (2011) Real-time inflation forecast densities from ensemble Phillips curves. *The North American Journal of Economics and Finance* **22**(1), 77–87.
- [21] Golub, G.H. and Van Loan, C.F. (2012) *Matrix computations (Vol. 3)*. JHU Press.

- [22] Golyandina, N., Nekrutkin, V.V., and Zhigljavski, A.A. (2001) *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman Hall/CRC Press.
- [23] Gray, A.G. and Thomson, P.J. (2002) On a family of finite moving average trend filters for the ends of series. *Journal of Forecasting* **21**(2), 125–149.
- [24] Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- [25] Harvey, A.C. and Delle Monache, D. (2009) Computing the mean square error of unobserved components extracted by misspecified time series models. *Journal of Economic Dynamics and Control* **33**(2), 283–295.
- [26] Henderson, R. (1916) Note on graduation by adjusted average. *Transactions of the actuarial society of America* **17**, 43–48.
- [27] Hodrick, R.J. and Prescott, E.C. (1997) Postwar US business cycles: an empirical investigation. *Journal of Money, credit, and Banking*, 1–16.
- [28] Lin, W., Huang, J.Z. and McElroy, T. (2019) Time series seasonal adjustment using regularized singular value decomposition. *Journal of Business & Economic Statistics*, published online, 1–38.
- [29] Lütkepohl, H. (2007) *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.
- [30] Maravall, A. and Pérez, D. (2012) Applying and interpreting model-based seasonal adjustment – the Euro-Area industrial production series. In *Economic Time Series: Modeling and Seasonality*, eds. Bell, W., Holan, S., McElroy, T. CRC Press. Boca Raton, FL.
- [31] McElroy, T. (2008a) Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* **24**(4), 988–1009.
- [32] McElroy, T. (2008b) Exact formulas for the Hodrick-Prescott filter. *The Econometrics Journal* **11**(1), 209–217.
- [33] McElroy, T. (2017) Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business & Economic Statistics* **35**(4), 611–625.

- [34] McElroy, T. and Findley, D. (2015) Fitting constrained vector autoregression models. In *Empirical Economic and Financial Research – Theory, Methods, and Practice*, eds. Beran, J., Feng, Y., and Hebbel, H. New York, Springer, 451–470.
- [35] McElroy, T. and McCracken, M. (2017) Multistep ahead forecasting of vector time series. *Econometric Reviews* **36**(5), 495–513.
- [36] McElroy, T. and Trimbur, T. (2015) Signal extraction for non-stationary multivariate time series with illustrations for trend inflation. *Journal of Time Series Analysis* **36**(2), 209–227.
- [37] McElroy, T. and Wildi, M. (2010) Signal extraction revision variances as a goodness-of-fit measure. *Journal of Time Series Econometrics* **2**(1).
- [38] Mise, E., Kim, T.H. and Newbold, P., 2005. On suboptimality of the HodrickPrescott filter at time series endpoints. *Journal of Macroeconomics* **27**(1), 53–67.
- [39] Musgrave, J. (1964) A set of end weights to end all end weights. Unpublished Working Paper. US Census Bureau.
- [40] Pollock, D.S.G. (2007) WienerKolmogorov filtering, frequency-selective filtering, and polynomial regression. *Econometric Theory* **23**(1), 71–88.
- [41] Pollock, D.S.G. (2009) Realisations of finite-sample frequency-selective filters. *Journal of Statistical Planning and Inference* **139**(4), 1541–1558.
- [42] Proietti, T. (2005) Forecasting and signal extraction with misspecified models. *Journal of Forecasting* **24**(8), 539–556.
- [43] Proietti, T. and Luati, A. (2008) Real time estimation in local polynomial regression, with application to trend-cycle analysis. *The Annals of Applied Statistics* **2**(4), 1523–1553.
- [44] Roy, A., McElroy, T., and Linton, P. (2019) Constrained estimation of causal invertible VARMA models. *Statistica Sinica* **29**, 455–478.
- [45] Schettlinger, K., Fried, R. and Gather, U. (2010) Real-time signal processing by adaptive repeated median filters. *International Journal of Adaptive Control and Signal Processing* **24**(5), 346–362.

- [46] Schorfheide, F. (2005) VAR forecasting under misspecification. *Journal of Econometrics* **128**(1), 99–136.
- [47] Schorfheide, F. and Song, D. (2015) Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics* **33**(3), 366–380.
- [48] Tiller, R. (2012) Frequency domain analysis of seasonal adjustment filters applied to periodic Labor Force Survey series. In *Economic Time Series: Modeling and Seasonality*, eds. Bell, W., Holan, S., McElroy, T. CRC Press. Boca Raton, FL.
- [49] Wildi, M. (2008) *Real-Time Signal Extraction: Beyond Maximum Likelihood Principles*. Berlin: Springer.
- [50] Wildi, M. (2018) Real-time trend extraction and seasonal adjustment. In *Handbook on Seasonal Adjustment*, eds. Mazzi, G., Ladiray, D., 415–449. Luxembourg: Publication Office of the European Union.
- [51] Wildi, M. and McElroy, T. (2016) Optimal real-time filters for linear prediction problems. *Journal of Time Series Econometrics* **8** (2), 155–192.
- [52] Wildi, M. and McElroy, T. (2019) The trilemma between accuracy, timeliness and smoothness in real-time signal extraction. *International Journal of Forecasting* **35** (3), 1072–1084.
- [53] Wildi, M. and Schips, B. (2004) Signal extraction: How (in) efficient are model-based approaches? an empirical study based on tramo/seats and census x-12-arima. KOF Working Papers, 96.