The Multivariate Linear Prediction Problem: Model-Based and Direct Filtering Solutions

Tucker S. McElroy[*†] and Marc Wildi[‡]

Abstract

Numerous contexts in macroeconomics, finance, and quality control require real-time estimation of trends, turning points, and anomalies. We formulate the real-time signal extraction problem as a multivariate linear prediction problem, present the optimal solution in terms of a known model, and propose multivariate direct filter analysis to address the more typical situation where the process' model is unknown. We show 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, and apply the methodology to petroleum, capitalization, and construction data.

Keywords. Frequency Domain, Seasonality, Time Series, Trends.

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

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

[*]Research and Methodology Directorate, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

[†]This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not necessarily those of the U.S. Census Bureau.

[‡]IDP, Zurich University of Applied Sciences, Rosenstrasse 8, 8401 Winterthur, Switzerland, marc.wildi@zhaw.ch

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.

However, these signals are not directly observable at time present, because in each case their definition involves all 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 processes 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, and useful, than historical signal extraction. The difficulty lies in the uncertainty about the future, which is transmitted unto the signal extraction estimates themselves.

There is a considerable body of literature addressing signal extraction (see below), but 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. Although an LPP can then be solved, the solution depends on knowing something about the dynamics in the time series process. The most venerable approach to understanding these dynamics is to posit a time series model, and fit this model to the observed data. 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).

MBA for forecasting problems is reviewed in Brockwell and Davis (1991), and the case of signal extraction problems is discussed in Bell and Martin (2004), as well as McElroy (2008). The multivariate case is discussed in McElroy and McCracken (2017) and McElroy and Trimbur (2015), for forecasting and signal extraction respectively; also see Harvey (1989) for a state space formulation.

Alexandrov et al. (2012) reviews trend filtering methods, while Dagum and Luati (2012), Maravall and Pérez (2012), Tiller (2012), and McElroy (2017) provide discussions of application to seasonal adjustment. However, there is still lacking a general multivariate solution to the LPP, which this article firstly provides.

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.

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.

These topics have been treated in Wildi and McElroy (2016, 2018) in the case of univariate time series, which discuss the basic DFA and the capabilities for customization. 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 and additional figures in the Supplementary material.

2 Background and Framework

Let $\{X_t\}$ be an 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 spectral density function (sdf) is a Hermitian matrix-valued function of $\omega \in [-\pi, \pi]$, defined as the Fourier Transform (FT) of the acf:

$$F(\omega) = \sum_{h \in \mathbb{Z}} \Gamma(h)\, z^h,$$

where we use the shorthand $z = e^{-i\omega}$. Given a bounded sdf (i.e., each F_{jk}, the (j,k)th entry of F, has bounded modulus as a function of ω), the acf can be recovered via inverse FT:

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

which uses the bracket notation to define the integral of a function (of ω) multiplied by $e^{i\omega h} = z^{-h}$, and the whole divided by 2π.

Datasets are typically available as a finite set of contiguous regular measurements, denoted $\{x_1, x_2, \ldots, x_T\}$, where T is the length of sample. The data is viewed as a realization of the corresponding random vectors $\{X_1, X_2, \ldots, X_T\}$, or alternatively as a time window of the sample path $\{x_t\}$ corresponding to times $1, 2, \ldots, T$. Applying the vec operator to such a sample yields the full vector $\underline{X}$, which is given by

$$\underline{X} = \mathrm{vec}[X_1, X_2, \ldots, X_T].$$

The covariance matrix of this NT-dimensional random vector, in the stationary case, is block Toeplitz, Each block is $N \times N$, and the stth such block, for $1 \le s, t \le T$, is given by $\Gamma(s-t)$. Also, from the sample we can compute the *Discrete Fourier Transform* (DFT) via

$$\widetilde{X}(\omega) = T^{-1/2} \sum_{t=1}^{T} z^t\, X_t. \tag{2}$$

This can be computed for any $\omega \in [-\pi, \pi]$, though for applications we restrict to the Fourier frequencies.

Definition 1 Given integer T, the Fourier frequencies are a set of T numbers in $[-\pi, \pi]$ of the form $\omega_j = 2\pi j/T$ for $-[T/2] \le j \le [T/2]$ (when T is odd) and $-[T/2] \le j \le [T/2] - 1$ (when T is even).

Remark 1 In the case that T is odd, there exists m such that $T = 2m + 1$, and in this case $-m \le j \le m$. In the case that T is even, then $T = 2m$ for some m, and $-m \le j \le m - 1$. Clearly, $m = [T/2]$ in either case.

The Fourier frequencies form the basis for a transformation of the time-domain sample $\underline{X}$ to the frequency-domain. By restricting the DFT to Fourier frequencies, we obtain a linear transformation from the $T \times N$ matrix of the sample to a $T \times N$ matrix of DFTs. To show this result, let

$$\mathcal{X} = [X_1, X_2, \ldots, X_T],$$

so that $\mathrm{vec}[\mathcal{X}] = \underline{X}$. Similarly, denote the matrix of DFTs by $\widetilde{\mathcal{X}}$, with jth column ($1 \leq j \leq T$) given by $\widetilde{X}(\omega_{j-[T/2]-1})$. In this way, the matrix of DFTs begins with $\widetilde{X}(\omega_{-[T/2]})$ in the first column, and proceeds to $\widetilde{X}(\omega_{T-[T/2]-1})$, with frequency corresponding to either $[T/2]$ or $[T/2]-1$ depending on whether T is odd or even. Letting C denote the $T \times T$ linear transformation such that $\widetilde{\mathcal{X}}' = C\,\mathcal{X}'$, we see that

$$C_{jt} = T^{-1/2}\,\exp\{-i\,2\pi t\,(j - [T/2] - 1)/T\},$$

for $1 \leq j, t \leq T$. This follows directly from (2). Moreover, the original sample can be recovered from the DFT matrix by applying C^{-1}, which equals the conjugate transpose (i.e., C is unitary).

The multivariate *periodogram* is defined to be the rank one Hermitian matrix

$$I(\omega) = \widetilde{X}(\omega)\,\widetilde{X}(\omega)^{*}. \tag{3}$$

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:

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

This is easily verified using the definition of sample autocovariance

$$\widehat{\Gamma}(h) = T^{-1}\sum_{t=1}^{T-h} X_{t+h}\,X_t'$$

for $h \geq 0$, and with $\widehat{\Gamma}(h) = \widehat{\Gamma}(-h)'$ for $h < 0$. Conversely, the periodogram is the FT of the sample autocovariances:

$$I(\omega) = \sum_{|h|<T} \widehat{\Gamma}(h)\,e^{-i\omega h}. \tag{5}$$

The lag operator on a time series is denoted L, and is defined via the action

$$L\,X_t = X_{t-1}.$$

Powers of L are defined analogously, with $L^0 = 1$ (an operator identity) and negative powers yielding forward time shifts, i.e., leads. Matrix polynomials of L yield new operators that act upon a time series using the linearity principle. Thus, if $A(L) = \sum_{j=0}^{p} a(j)\,L^j$ for $N \times N$ matrices $a(j)$, then

$$A(L)\,X_t = \sum_{j=0}^{p} a(j)\,X_{t-j}.$$

The latent dynamics of $\{X_t\}$ can be revealed through the application of a multivariate linear filter $\Psi(L) = \sum_{j\in\mathbb{Z}} \psi(j)\,L^j$. The properties of a filter can be studied by setting $L = z$, thereby obtaining the frequency response function (frf): $\Psi(e^{-i\omega}) = \sum_{j\in\mathbb{Z}} \psi(j)\,e^{-i\omega j}$. Another quantity of interest is the derivative of a filter, defined via $\partial\Psi(L) = \sum_{j\in\mathbb{Z}} j\,\psi(j)\,L^{j-1}$. An example of $\Psi(L)$ is provided

by a low-pass filter that extracts long-run dynamics, such as trends and business-cycles. In general the extracted signal is the output of the filter, i.e.,

$$Y_t = \Psi(L)\, X_t = \sum_{j\in\mathbb{Z}} \psi(j)\, X_{t-j}. \tag{6}$$

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 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. We seek a causal $\widehat{\Psi}(L) = \sum_{j\geq 0} \widehat{\psi}(j)\, L^j$ that approximates $\Psi(L)$ on the time series of interest, i.e., the filter error

$$E_t = \Psi(L)X_t - \widehat{\Psi}(L)X_t \tag{7}$$

should be stationary, mean zero, and have small variance. Because E_t is an N-vector, we can design a criterion that involves minimizing the trace of its covariance matrix, for example. The quest for $\widehat{\Psi}(L)$, such that $\mathrm{Var}[E_t]$ is minimal, is called the *linear prediction problem* (LPP).

A model-based (MB) approach to the problem proceeds as follows: we can compute the optimal $\widehat{\Psi}(L)$ analytically, given knowledge of $\Psi(L)$ 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.) The univariate solution was given in Wildi and McElroy (2016), and the multivariate solution is described below. These MB solutions to the LPP yield a formula for $\widehat{\Psi}(L)$, which can then be applied to generate real-time signals.

A deficiency with the MB approach is mis-specification: we must have the exact Wold decomposition of the process. Multivariate direct filter analysis (MDFA) instead foregoes knowledge of the exact dynamics, and attempts to minimize $\mathrm{Var}[E_t]$ with respect to the unknown coefficients of $\widehat{\Psi}(L)$. One can take $\widehat{\Psi}(L)$ belonging to some class of causal filters, and parametrize that class via θ; then minimize the criterion with respect to θ. In order to do this, one must estimate the process' spectral density, and it is adequate from a theoretical perspective to use the periodogram, as shown below. This is the basic MDFA solution to an LPP.

We mention two extensions of this basic MDFA. First, it may be of interest to constrain the solution $\widehat{\Psi}(L)$ in various ways – this can be done by restricting the class of causal filters. For instance, it may be of interest to ensure that $\widehat{\Psi}(L)$ and $\Psi(L)$ 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 2 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 $\Psi(L)$ acting on a given observed time series $\{X_t\}$ if and only if we can write for all integers t

$$Y_t = \Psi(L)\, X_t.$$

We say that $\{Y_t\}$ is a **scalar target** if $\Psi(L)$ is a $1 \times N$-dimensional filter.

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 one of the component series h steps ahead, where $h \geq 1$ is the given *forecast lead*. Here, suppose that the series of interest is the first component, so that

$$Y_t = X_{t+h}$$

for all $t \in \mathbb{Z}$. This target corresponds to $\Psi(L) = L^{-h}\, 1_N$, where 1_N is the identity matrix of dimension N. Thus, $\psi(\ell)$ is a $N \times N$ matrix, each of which are zero except $\psi(-h)$, which is given by 1_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 $\Psi(z)$ 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. For scalar target, we have

$$\Psi(z) = \chi_{[-\mu,\mu]}(\omega)\, 1_N$$

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

$$\psi(\ell) = \frac{\sin(\ell\mu)}{\pi\ell}\, 1_N$$

for $\ell \neq 0$ and $\psi(0) = \mu/\pi\, e_1'$.

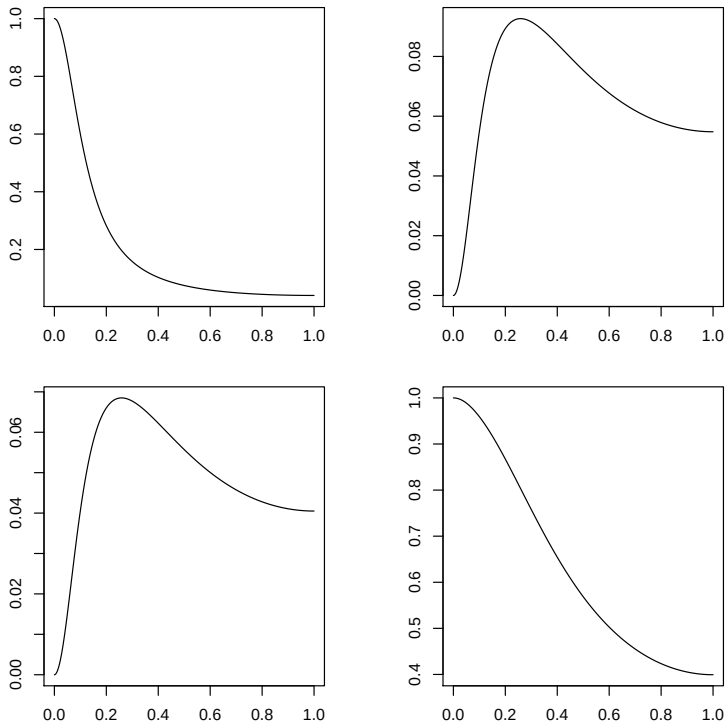

Figure 1: Bivariate trend frequency response functions for Petroleum data, with input series by column, and output series by row.

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 $\{\mu_t\}$ and a white noise irregular $\{\iota_t\}$, such that $X_t = \mu_t + \iota_t$. Both the multivariate trend and irregular are driven by independent white noise processes, with respective covariance matrices Σ_μ and Σ_ι, and the frf for the optimal trend extraction filter (McElroy and Trimbur, 2015) is

$$\Psi(e^{-i\omega}) = \Sigma_\mu \left[\Sigma_\mu + (2 - 2\,\cos(\omega))\,\Sigma_\iota\right]^{-1}.$$

An example – based on a bivariate LLM fitted to Petroleum data discussed below – yields the frfs plotted in Figure 1. Note that the upper left panel has a low-pass shape, but the lower right panel passes a lot of noise. (Both cross-gains, in the off-diagonal panels, offer a small contribution.)

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 $\{\mu_t\}$ is an integrated random walk, i.e., $(1-L)^2\mu_t$ is white noise of covariance matrix Σ_μ. Then the frf for the optimal trend extraction filter – which also coincides with the multivariate Hodrick-Prescott filter (cf. McElroy and Trimbur, 2015) – is given by

$$\Psi(e^{-i\omega}) = \Sigma_\mu \left[\Sigma_\mu + (2 - 2\,\cos(\omega))^2\,\Sigma_\iota\right]^{-1}.$$

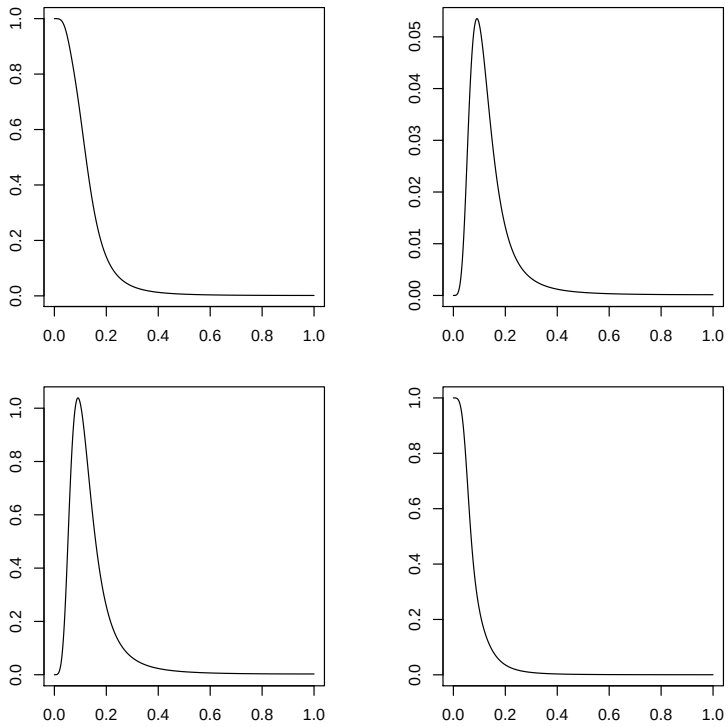

Figure 2: Bivariate trend frequency response functions for Non-Defense Capitalization data, with input series by column, and output series by row.

The chief difference with the frf of the LLM is that the sinusoidal factor is now squared. Fitting this model to the bivariate time series of Non-Defense Capitalization data (Shipments and New Orders, March 992 through July 2016, in seasonally adjusted form) yields the frfs plotted in Figure 2. There is much more high-frequency content filtered out by this filter, as compared with the LLM trend filter (cf. Figure 1); also notable, is the substantial cross-gain at business cycle frequencies in the lower left panel.

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 where $\delta(L) = (1-L)(1-L^{12})$, we obtain an integrated random walk trend component $\{\mu_t\}$ (identical to the trend discussed in Example 4) and six atomic seasonal components that combine into a single seasonal component $\{\xi_t\}$ with differencing operator $U(L) = 1+L+L^2+\ldots+L^{11}$, along with the irregular $\{\iota_t\}$. Six separate covariance matrices govern the dynamics of the seasonal component, allowing for different degrees of smoothness at each of the six seasonal frequencies. Fitting this model to the Housing Starts data (discussed below) results in filter frfs with nuanced behavior – see Figure 3 for the seasonal adjustment filters. The troughs at seasonal frequencies are typical, but note that breadth varies by series; moreover, the spectral

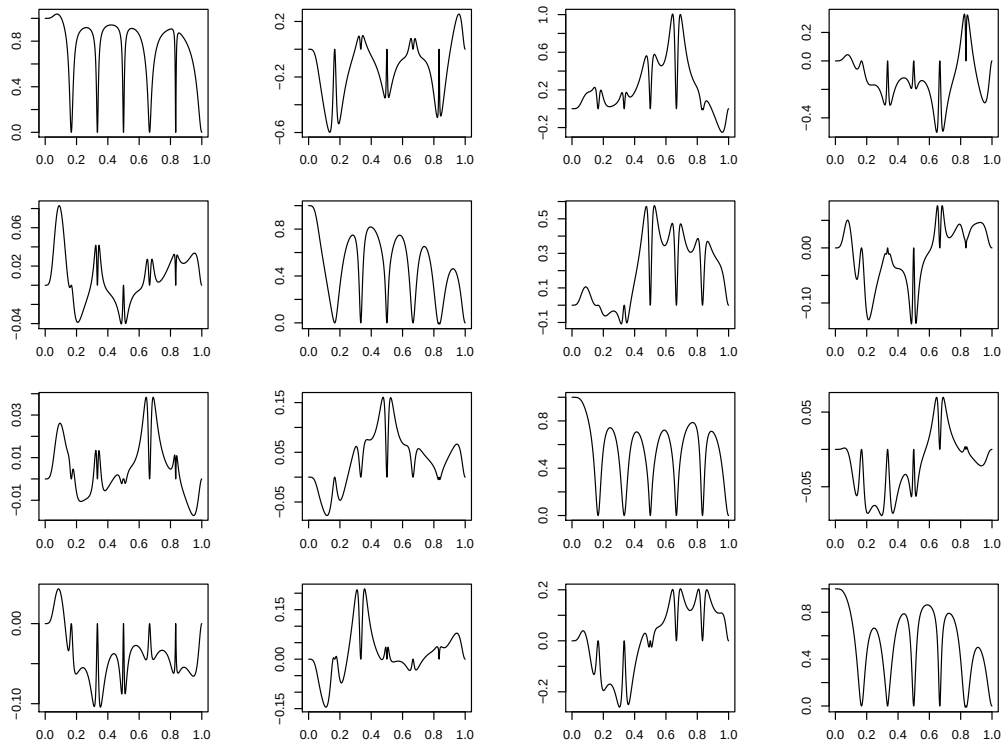

Figure 3: Quadvariate seasonal adjustment frequency response functions for Housing Starts data, with input series by column, and output series by row.

content in some of the cross-gains is pronounced (e.g., the third panel of the first row, representing the impact of the NorthEast sector upon the seasonal adjustment filter of the South sector).

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 X_t.

Definition 3 The **Linear Prediction Problem** (LPP) seeks a linear estimate of the form (6) such that the filter error (7) has mean zero such that the determinant of the filter error variance $\mathrm{Var}[E_t]$ 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 $X_t = \Theta(L)\,\epsilon_t$, where $\{\epsilon_t\}$ is WN(Σ) and $\Theta(L) = \sum_{\ell\in\mathbb{Z}} \theta(\ell)\,L^\ell$. When $\theta(\ell) = 0$ for all $\ell < 0$, the process is called *causal*. For any power series, we introduce the notation $[\Theta(L)]_a^b = \sum_{j=a}^b \theta(j)\,L^j$. With these preliminaries, we can state the solution to the LPP.

Proposition 1 *Suppose that $\{X_t\}$ is mean zero and weakly stationary with causal Wold decomposition expressed as $X_t = \Theta(L)\,\epsilon_t$. Then the solution to the LPP posed by a target $Y_t = \Psi(L)\,X_t$ is given by*

$$\widehat{\Psi}(L) = \sum_{\ell \geq 0} \psi(\ell)\,L^{\ell} + \sum_{\ell < 0} \psi(\ell)\,[\Theta(L)]_{-\ell}^{\infty}\,L^{\ell}\,\Theta(L)^{-1}. \tag{8}$$

Moreover, the MSE corresponding to this solution is given by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{\ell > 0} \psi(-\ell)\,z^{-\ell}[\Theta(z)]_0^{\ell-1}\,\Sigma\,[\Theta(z)]_0^{\ell-1*} \sum_{\ell > 0} \psi(-\ell)\,z^{\ell}\,d\omega. \tag{9}$$

Remark 2 The formula (9) gives us a lower bound on the determinant of the MSE when we use sub-optimal proxies for $\widehat{\Psi}(L)$.

As indicated by Remark 2, the result of Proposition 1 is chiefly useful when we know $\Theta(L)$. However, this is rarely the case in practice: a 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 $\Theta(L) = (1 - \Phi\,L)^{-1}$. Because for $\ell < 0$

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

we find that (8) reduces to

$$\widehat{\Psi}(L) = \sum_{\ell \geq 0} \psi(\ell)\,L^{\ell} + \sum_{\ell < 0} \psi(\ell)\,\Phi^{-\ell}. \tag{10}$$

The second term in this expression we denote by $A_{\Psi}(\Phi)$. Hence, the optimal concurrent filter is determined by applying the filter to past data and modifying the present weight $\psi(0)$ by adding the quantity $A_{\Psi}(\Phi)$. In the case of h-step ahead forecasting of the first time series (Example 1), $\widehat{\Psi}(L) = A_{\Psi}(\Phi) = \Phi^h$. 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 (2018).)

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 $\widehat{\Psi}(B)$ 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 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 $\Psi(L)$, we use the notation

$$\mathcal{G} = \{\widehat{\Psi}_\vartheta(L) : \vartheta \text{ belongs to a parameter space}\}. \tag{11}$$

First suppose that $\{X_t\}$ is weakly stationary with mean zero and spectral density F. The real-time estimation error is given in (7), which has mean zero and $N \times N$ variance matrix

$$\mathbb{E}[E_t\, E_t'] = \langle \left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]\, F\, \left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]^* \rangle_0. \tag{12}$$

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

$$D_\Psi(\vartheta, G) = \langle \left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]\, G\, \left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]^* \rangle_0. \tag{13}$$

In the following development, setting $G = F$ yields an ideal criterion based on the process, whereas setting $G = I$ (the periodogram) yields an empirical criterion, providing estimates that we can compute from data. Taking the determinant of (13) yields the MDFA criterion function. Given a filter class $\mathcal{G}$, the best possible concurrent filter is given by $\Gamma_{\vartheta(F)}$, 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

$$\vartheta' = \left[\widehat{\psi}(0), \widehat{\psi}(1), \ldots, \widehat{\psi}(q-1)\right]. \tag{14}$$

So ϑ is a $qN \times N$ dimensional matrix. Then the criterion (13) can be rewritten as

$$D_\Psi(\vartheta, G) = \vartheta'\, B\, \vartheta - \vartheta'\, b - b'\, \vartheta + \langle \Psi(z)\, G\, \Psi(z)^* \rangle_0, \tag{15}$$

where

$$b' = \left[\langle \Psi(z)\, G\rangle_0, \langle \Psi(z)\, G\rangle_1, \ldots, \langle \Psi(z)\, G\rangle_{q-1}\right], \tag{16}$$

and B is a block matrix, where the jkth $N \times N$ block of is $\langle G\rangle_{k-j}$ for $1 \le j, k \le q$. (Because G is Hermitian, $\langle G\rangle_{k-j}$ is real, and it follows that b is real as well.)

Proposition 2 *The minimizer of the MDFA criterion given by the determinant of (13), with respect to $\mathcal{G}$ consisting of all length q concurrent filters, is*

$$\vartheta(G) = B^{-1}\, b,$$

where the jkth block of B is $\langle G \rangle_{k-j}$, and b is given by (16). The minimal value is the determinant of

$$\langle \Psi(z)\, G\, \Psi(z)^* \rangle_0 - b'\, B^{-1}\, b. \tag{17}$$

Remark 3 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 b, given by (16), and we can proceed by approximating the integrals over a Riemann mesh corresponding to Fourier frequencies.

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

$$\vartheta = \text{vec}[\widehat{\psi}(0)', \widehat{\psi}(1)', \ldots, \widehat{\psi}(q-1)'].$$

With $\Psi(L) = L^{-1}$ from (13) we have

$$\begin{aligned} D_\Psi(\vartheta, G) &= \langle \left[z^{-1} 1_N - \widehat{\Psi}_\vartheta(z) \right]\, G \left[z^{-1} 1_N - \widehat{\Psi}_\vartheta(z) \right]^* \rangle_0 \\ &= \langle \left[1_N - \sum_{\ell=0}^{q-1} \widehat{\psi}(\ell)\, z^{\ell+1} \right]\, G \left[1_N - \sum_{\ell=0}^{q-1} \widehat{\psi}(\ell)\, z^{\ell+1} \right]^* \rangle_0 \\ &= \langle G \rangle_0 - 2\, \vartheta'\, \langle G \rangle_{1:q} + \vartheta'\, \langle G \rangle_{0:(q-1),0:(q-1)}\, \vartheta. \end{aligned}$$

Hence the optimizer is

$$\vartheta(G) = \langle G \rangle^{-1}_{0:(q-1),0:(q-1)}\, \langle G \rangle_{1:q},$$

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 $G = I$ the LPF can be constructed directly from the sample.

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

$$\varphi' = [\psi(0) + A_\Psi(\Phi), \psi(1), \ldots, \psi(q-1)].$$

This is an approximate solution to the system $\vartheta' B = b'$, because $\varphi' B$ has $j+1$th component, for $0 \le j \le q-1$, equal to

$$\sum_{\ell=0}^{q-1} \psi(\ell) \langle G \rangle_{j-\ell} + A_\Psi(\Phi)\,\Gamma(j).$$

Noting that

$$A_\Psi(\Phi)\,\Gamma(j) = \sum_{\ell<0} \psi(-\ell)\,\Phi^{-\ell}\,\Gamma(j) = \sum_{\ell<0} \psi(-\ell)\,\Gamma(j-\ell),$$

because for a VAR(1) process $\Gamma(h) = \Phi^h\,\Gamma(0)$ when $h \ge 0$, we see that component $j+1$ of $\varphi' B$ is

$$\sum_{\ell \le q-1} \psi(\ell)\,\Gamma(j-\ell) = [b']_{j+1} - \sum_{\ell \ge q} \psi(\ell)\,\Gamma(j-\ell).$$

As $q \to \infty$ the error term vanishes (for each j), indicating that $\varphi' B \approx b'$, or $\vartheta \approx \varphi$.

To compute the quantities given in Proposition 2, and more generally to compute the MDFA criterion (13), 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 $E_t = Y_t - \widehat{Y}_t$ has variance given by (12), the sample variance is

$$T^{-1} \sum_{t=1}^{T} E_t\,E_t' = T^{-1} \sum_{j=1-[T/2]-1}^{T-[T/2]-1} I_E(\omega_j),$$

where I_E is the periodogram of the filter errors. This equality is a discrete version of the Plancherel identity; the right hand side (with I_X the periodogram of the process) is approximated by

$$T^{-1} \sum_{j=1-[T/2]-1}^{T-[T/2]-1} \left[\Psi(e^{-i\omega_j}) - \widehat{\Psi}(e^{-i\omega_j})\right]\, I_X(\omega_j) \left[\Psi(e^{i\omega_j}) - \widehat{\Psi}(e^{i\omega_j})\right]'.$$

This is exactly the criterion $D_\Psi(\vartheta, I_X)$ of (13) 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

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

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

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

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

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

4.2 Constrained MDFA

Various constraints upon the concurrent filter can be envisioned, and imposing such strictures results in a constrained MDFA. Writing $\Delta(B) = \Psi(B) - \widehat{\Psi}(B)$ as the discrepancy filter, we see from (7) that $\mathbb{E}[E_t] = \Delta(B)\,\mathbb{E}[X_t]$; by Definition 3, we require that $\mathbb{E}[E_t] = 0$ for any LPP. If $\mathbb{E}[X_t] = 0$ then this condition is always satisfied, but with nonzero means additional constraints on $\Delta(B)$ must be imposed, which implicitly amount to constraints on $\widehat{\Psi}(B)$. The following results are well-known (Brockwell and Davis, 1991): if $\mathbb{E}[X_t]$ is constant but nonzero, then we require $\Delta(1) = 0$. If $\mathbb{E}[X_t]$ 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:

$$\begin{aligned}
\text{LC} &: \Delta(1) = 0 \quad \text{or} \quad \Psi(1) = \widehat{\Psi}(1) \\
\text{TSC} &: \partial\Delta(1) = 0 \quad \text{or} \quad \partial\Psi(1) = \partial\widehat{\Psi}(1) \\
\text{LTSC} &: \Delta(1) = 0,\ \partial\Delta(1) = 0 \quad \text{or} \quad \Psi(1) = \widehat{\Psi}(1),\ \partial\Psi(1) = \partial\widehat{\Psi}(1).
\end{aligned}$$

In the case of concurrent filters of form (14), LC is accomplished by demanding that $\sum_{j=0}^{q-1} \widehat{\psi}(j) = \Psi(1)$. More generally, we consider linear constraints formulated via

$$\vartheta = R\,\varphi + Q, \tag{18}$$

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 vec$[\vartheta']$, but is sufficient to describe LC, TSC, and LTSC.

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

$$\widehat{\psi}(0) = \Psi(1) - \sum_{j=1}^{q-1} \widehat{\psi}(j). \tag{19}$$

Hence $\varphi' = [\widehat{\psi}(1), \widehat{\psi}(2), \ldots, \widehat{\psi}(q-1)]$ and

$$R = \begin{bmatrix} -1 & \ldots & -1 \\ 1 & 0 & 0 \\ \vdots & \ddots & \vdots \\ 0 & 0 & 1 \end{bmatrix} \otimes 1_N \qquad Q = \begin{bmatrix} \Psi(1) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

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 $\varphi' = [\widehat{\psi}(0), \widehat{\psi}(2), \ldots, \widehat{\psi}(q-1)]$ and

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

Level and Time Shift Constraint (LTSC). Take the Time Shift constraint formula for $\widehat{\psi}(1)$, and plug this into (19), 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 $\varphi' = [\widehat{\psi}(2), \ldots, \widehat{\psi}(q-1)]$ and

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

More generally, we can envision an LPP involving M linear constraints on each scalar filter in ϑ, taking the form $A = [J \otimes 1_N]\,\vartheta$, where J is $M \times q$ dimensional ($M < q$) and A is $NM \times N$ dimensional. (The LC, TSC, and LTSC examples all have this form.) In order to express this constraint in the form (18), we use the Q-R decomposition (Golub and Van Loan, 1996) of J, writing $J = C\,G\,\Pi$ 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 permuation 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 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

$$\left[G_1^{-1}\,C^{-1} \otimes 1_N\right]\, A = \left(\left[1_M,\, G_1^{-1}\,G_2\right]\,\Pi \otimes 1_N\right)\,\vartheta,$$

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

$$\begin{bmatrix} \overline{\vartheta} \\ \underline{\vartheta} \end{bmatrix} = (\Pi \otimes 1_N)\ \vartheta,$$

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

$$\overline{\vartheta} = \left[G_1^{-1}\, C^{-1} \otimes 1_N\right]\ A - \left[G_1^{-1}\, G_2 \otimes 1_N\right]\ \underline{\vartheta}.$$

Therefore we recognize the free variables $\varphi = \underline{\vartheta}$, and obtain R and Q in (18) via

$$R = \Pi^{-1} \begin{bmatrix} -G_1^{-1}\, G_2 \\ 1_r \end{bmatrix} \otimes 1_N$$

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

These formulas allow one to compute the form (18) 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 (13), with respect to $\mathcal{G}$ consists of all length q concurrent filters subject to linear constraints of the form (18), is*

$$\varphi = \left[R'\, B\, R\right]^{-1} R'\ (b - B\, Q)\,. \tag{20}$$

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

$$\langle \Psi(z)\, G\, \Psi(z)^* \rangle_0 - b'\, R\left[R'\, B\, R\right]^{-1} R'\, b + Q'\, B\, H\, Q - 2\, b'\, H\, Q. \tag{21}$$

For computation, we utilize the same approximations to B and b as discussed in the previous subsection, obtaining the constrained MDFA filter ϑ via (20) followed by (18).

4.3 Non-stationary MDFA

We here consider difference-stationary vector time series, which means there exists a scalar differencing polynomial $\delta(L)$ such that $\partial X_t = \delta(L) X_t$ is mean zero and covariance stationary. Examination of (7) indicates that the error process is not stationary unless we make certain assumptions about $\Delta(L) = \Psi(L) - \widehat{\Psi}(L)$. It is necessary that we can factor $\delta(L)$ from $\Delta(L)$, i.e., there exists $\widetilde{\Delta}(L)$ such that

$$\Delta(L) = \widetilde{\Delta}(L)\, \delta(L), \tag{22}$$

as otherwise we cannot guarantee that $\{E_t\}$ will be stationary. However, (22) is sufficient to guarantee that the filter error be stationary, because

$$E_t = \widetilde{\Delta}(L)\, \partial X_t$$

in such as case. We next discuss a set of filter constraints that guarantee (22), beginning with a lemma that discusses factorization of filters. We say a filter $\Psi(L)$ is absolutely convergent if $\sum_{j\in\mathbb{Z}} \|\psi(j)\| < \infty$ for a given matrix norm $\|\cdot\|$.

Proposition 4 *Any linear filter $\Psi(L)$ can be expressed as*

$$\Psi(L) = \Psi(\zeta) + (L-\zeta)\,\Psi^\sharp(L)$$

for any $\zeta \in \mathbb{C}$ such that $|\zeta| = 1$, and an absolutely convergent filter $\Psi^\sharp(L)$, so long as $\partial\Psi(L)$ is absolutely convergent. If in addition $\partial\partial\Psi(L) = \sum_{j\in\mathbb{Z}} j(j-1)\,\psi(j)\,L^j$ is absolutely convergent, then there also exists an absolutely convergent filter $\Psi^\flat(L)$ such that

$$\Psi(L) = \Psi(\zeta) + \partial\Psi(\zeta)\,(L-\zeta)\,\overline{\zeta} + (L-\zeta)^2\,\Psi^\flat(L).$$

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

Definition 4 For $\omega \in [-\pi, \pi]$, a filter $\Psi(L)$ 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 $\Psi(L)$ annihilates ω-noise of order 1 and $\partial\Psi(L)$ is absolutely convergent, then*

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

If a filter $\Psi(L)$ annihilate ω-noise of order 2, and $\partial\partial\Psi(L)$ is absolutely convergent, then

$$\Psi(L) = (L - e^{-i\omega})^2\,\Psi^\flat(L).$$

We can apply Corollary 1 to factor a noise-differencing polynomial $\delta^N(L)$ from $\Delta(L)$: for each ω such that the target filter $\Psi(L)$ annihilate ω-noise of order d, we impose the constraint that $\widehat{\Psi}(L)$ shall have the same property, and hence $(L - e^{-i\omega})^d$ 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

$$\Psi(L) = \prod_\ell (L - e^{-i\omega_\ell})^{d_\ell}\,\Psi^\natural(L) = \delta^N(L)\,\Psi^\star(L)$$

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

$$\Delta(L) = \left(\Psi^\star(L) - \widehat{\Psi}^\star(L)\right)\,\delta^N(L). \tag{23}$$

So if $\delta(L) = \delta^N(L)$, then (22) 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 (2008). Many signal extraction filters $\Psi(L)$ have the property that they annihilate ω-noise of the appropriate order, such that $\delta^N(L)$ can be factored; in addition, the noise filter $1_N - \Psi(L)$ has the same property with respect to the signal frequencies, i.e., $\delta^S(L)$ can be factored from $1_N - \Psi(L)$ in the same manner. Hence $1_N - \Psi(L) = \delta^S(L)\,\Psi^\diamond(L)$ for some factor $\Psi^\diamond(L)$, and imposing the same constraints on the concurrent filter yields

$$\Delta(L) = (1_N - \widehat{\Psi}(L)) - (1_N - \Psi(L)) = \left(\widehat{\Psi}^\diamond(L) - \Psi^\diamond(L)\right)\,\delta^S(L).$$

However, (23) 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 $\delta(L) = \delta^N(L)\,\delta^S(L)$ must be a factor. Therefore, $\widetilde{\Delta}(L) = (\widehat{\Psi}^\diamond(L) - \Psi^\diamond(L))/\delta^N(L)$, and (22) holds.

In summary, given a factorization of $\delta(z)$ into signal and noise differencing polynomials, the noise constraints and signal constraints on $\Psi(L)$ must also be imposed upon $\widehat{\Psi}(L)$, and this ensures that $\{E_t\}$ will be stationary with mean zero. If ω satisfies $\delta^N(e^{-i\omega}) = 0$, then we impose that $\widehat{\Psi}(L)$ 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 $\widehat{\Psi}(L)$ in the form $A = [J \otimes 1_N]\,\vartheta$ for filters of form (14), applying the methodology of the previous subsection. With these constraints in play, the formula (12) 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

$$D_\Psi(\vartheta, G) = \langle \widetilde{\Delta}(z)\,G\,\widetilde{\Delta}(z)^* \rangle_0 = \langle \left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]\,G\,|\delta(z)|^{-2}\left[\Psi(z) - \widehat{\Psi}_\vartheta(z)\right]^* \rangle_0. \tag{24}$$

The second expression in (24) utilizes (22), 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 ω for $z = e^{-i\omega}$, as the previous discussion guarantees. As a matter of convenience, given that the frequencies of singularity in $|\delta(z)|^{-2}$ 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 $\delta(z)$ is nonzero. Whereas the theoretical filter error MSE is given by $D_{\Psi,F}$, 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 $\delta(z)$, as such an omission only results in a loss of order T^{-1}. (The alternative is to compute the quantities $\widetilde{\Delta}(z)$ 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.

5.1 VAR(1) Specification and Trend Extraction

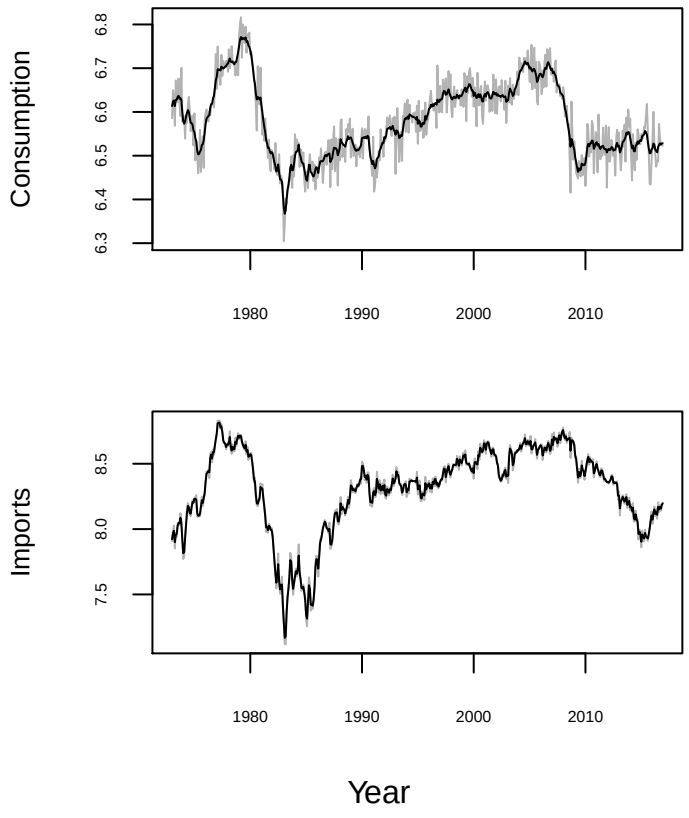

Figure 4: Bivariate Petroleum data (1973.1 through 2016.12) in grey, with black MB trends.

We suppose the true process is a VAR(1), and apply the MB trend filter defined in Example 3. The parameters of the filter have been obtained by fitting a LLM to a bivariate petroleum series: Industrial Petroleum Consumption and OPEC Oil Imports, Jan 1973 through December 2016, both seasonally adjusted, which are displayed in Figure 4. The MLEs for the model parameters are

$$\Sigma_\mu = \begin{bmatrix} 2.32 \cdot 10^{-4} & 5.04 \cdot 10^{-4} \\ 5.04 \cdot 10^{-4} & 34.73 \cdot 10^{-4} \end{bmatrix} \qquad \Sigma_\iota = \begin{bmatrix} 110.44 \cdot 10^{-5} & 7.17 \cdot 10^{-5} \\ 7.17 \cdot 10^{-5} & 128.57 \cdot 10^{-5} \end{bmatrix}.$$

As a result, the trend signal is quite a bit stronger in the second series (Imports), indicating that little smoothing is needed; this is why the trend for the second series closely tracks the original data (second panel of Figure 4).

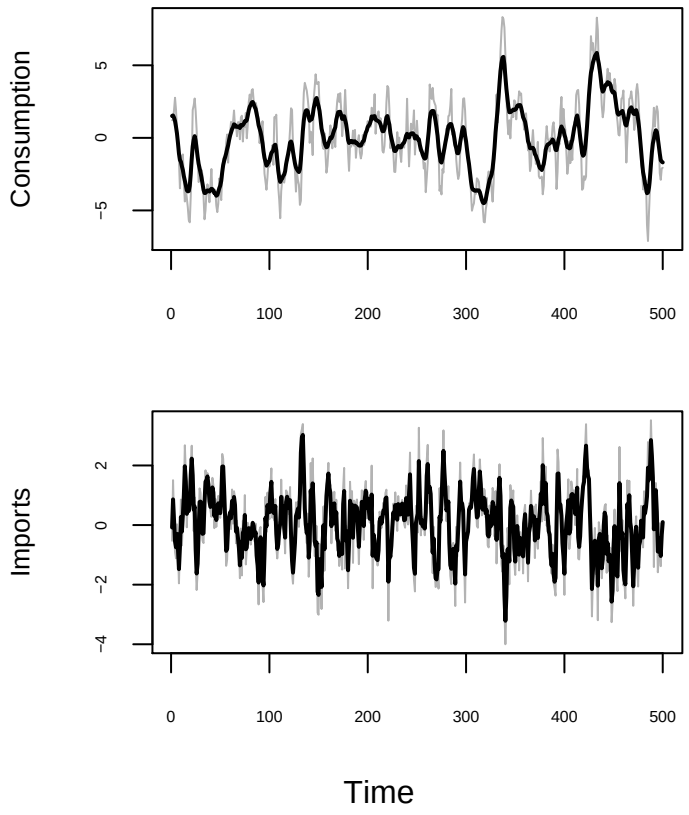

Figure 5: Bivariate trend filter applied to VAR(1) process (grey), with trends in black.

We seek to solve the corresponding trend extraction LPP. First, we can use the optimal solution (10) given in Illustration 1, using our knowledge that the VAR(1) is correctly specified. 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. The ideal trends $\Psi(B)X_t$ are produced by truncating the filter to length 4001 (it is symmetric, so the indices range between -2000 and 2000) and applying to a simulation of length 4500, only retaining the central 500 data points, as displayed in Figure 5. (In this way we can dispense with edge effects, and the extra 4000 observations are not used in the MDFA.) The black line of Figure 5 is the target, and we wish to use MDFA (setting $q = 30$) with various constraints (LC, TSC, LTSC) to obtain a good real-time estimate, comparing to the optimal solution given by implementing Illustration 1:

we find that

$$A_\Psi(\Phi) = \begin{bmatrix} 0.317 & 0.218 \\ -0.054 & 0.027 \end{bmatrix},$$

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 statistical error. When imposing a level constraint (LC and LTSC) there is a 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

$$\widehat{\Psi}(1) = \begin{bmatrix} 0.914 & 0.251 \\ -0.030 & 0.842 \end{bmatrix},$$

which is quite different from 1_2. On the other hand, the time shift constraint alone (TSC) has little impact on the performance of MDFA, because $\partial\widehat{\Psi}(1) \approx 0 \cdot 1_2$, i.e., the optimal filter already has this property of zero time shift.

Series	LPP Opt	MDFA Basic	MDFA LC	MDFA TSC	MDFA LTSC
1	.2669	.2668	.3093	.2701	.4452
2	.0234	.0232	.0246	.0234	.0285

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

We now extend the previous exercise by studying the LLM determined by the Petrol MLEs; 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 $\widehat{\Psi}(B)$ 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 6.

We apply MDFA in the manner described in Section 4.3, where $\delta^S(L) = 1 - L$ and $\delta^N(L) = 1$. 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 $\widehat{\Psi}(B)$. Results are displayed in Figure B.1 of the Appendix, with the target trend in black (these are the same black trend lines as displayed in Figure 6, but here shown without the underlying simulation) and the MDFA real-time

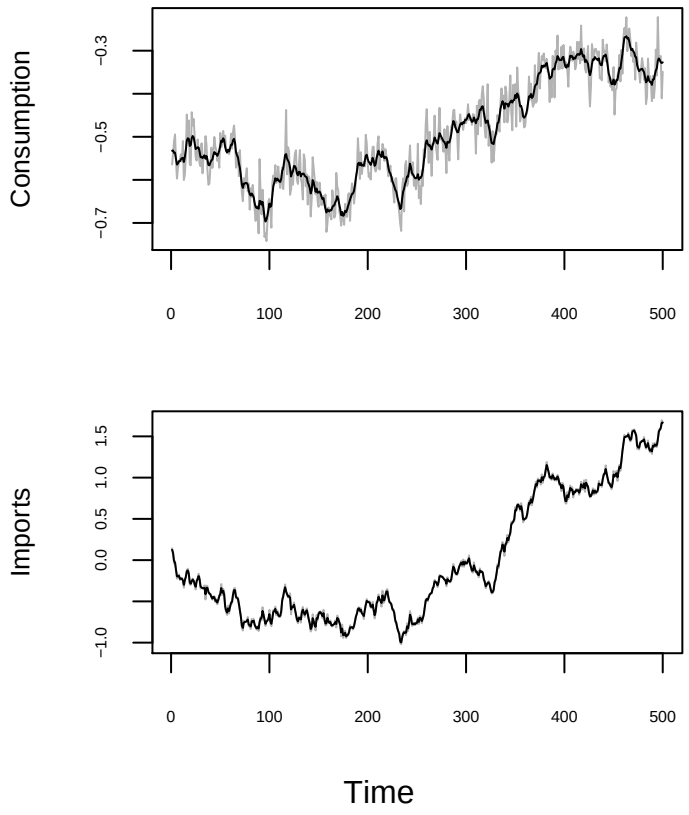

Figure 6: Bivariate trend filter applied to Null LLM process (grey), with trends in black.

	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}.)

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.

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

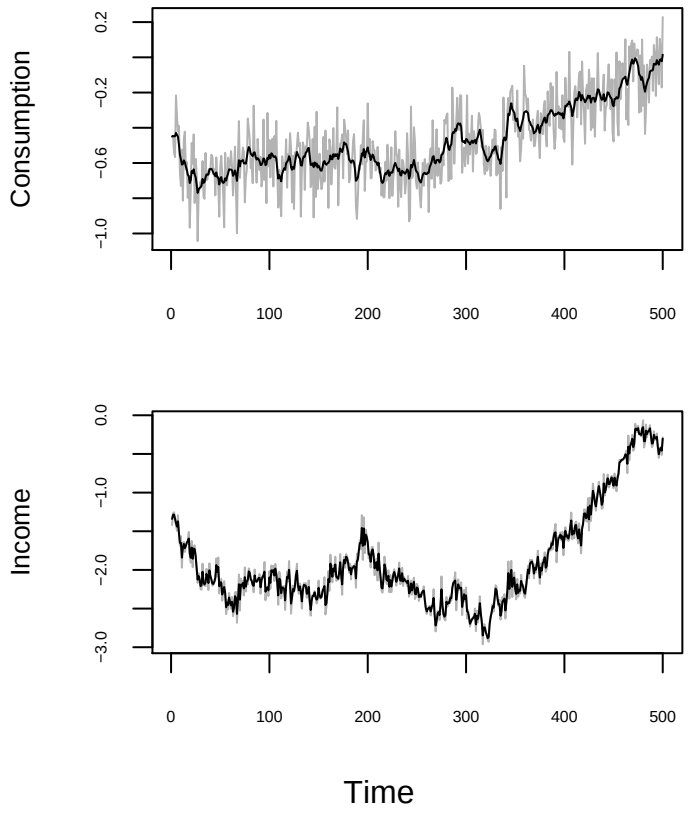

Figure 7: Bivariate trend filter applied to alternative LLM process (grey), with trends in black.

simulation – now the MB concurrent will do too little smoothing. The new irregular covariance matrix is

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

We refer to this as the alternative LLM process. The resulting simulation with trends is shown in Figure 7. Application of MDFA yields the trends plotted in Figure B.2 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 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 $\Psi(B)$. Taking the same MB filter as target, we compute the non-stationary MDFA solution to the LPP, and compare to the concurrent MB

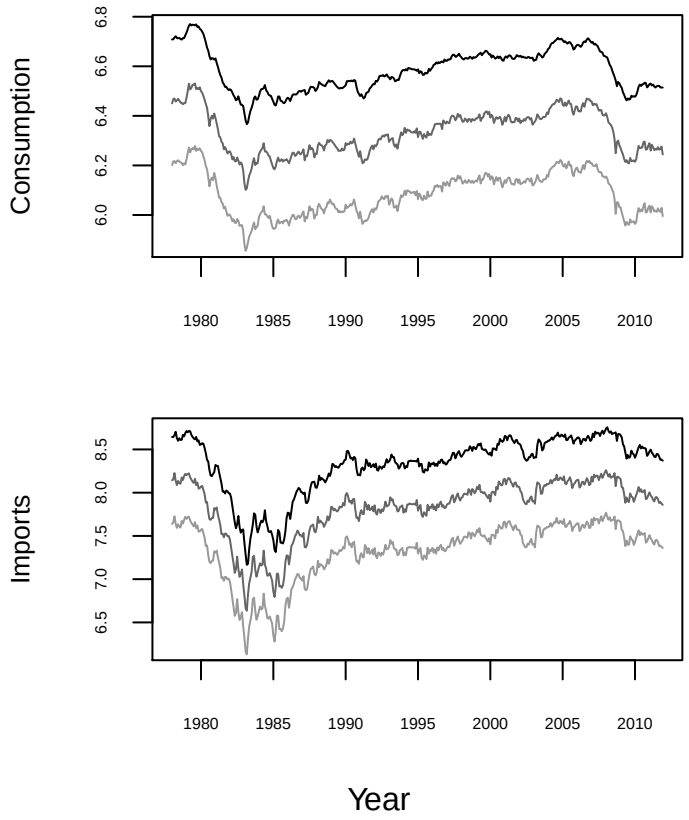

Figure 8: Petroleum data LLM bivariate trend output (black), with non-stationary MDFA trend (dark grey) and MB concurrent trend (light grey). The trend lines have been vertically staggered for easier visualization.

filter output. The resulting trends are plotted in Figure 8. The in-sample MSE is displayed in the final block of columns Table 2, showing a modest improvement for MDFA on the first series, and similar performance on the second series.

5.3 Housing Starts

Finally, we consider the quavariate time series of Housing Starts (South, West, NorthEast, Mid-West) January 1964 through December 2012, abbreviated as starts. The series is displayed in Figure 9, along with MB seasonal adjustments obtained from a fitted structural model. The MLEs are omitted (there are 8 matrices of dimension 4), but the impact of differing variances among the components and sectors can be discerned in the seasonal adjustment frf displayed in Figure 3.

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 $(1 - L)^2$ (i.e., a double root at frequency

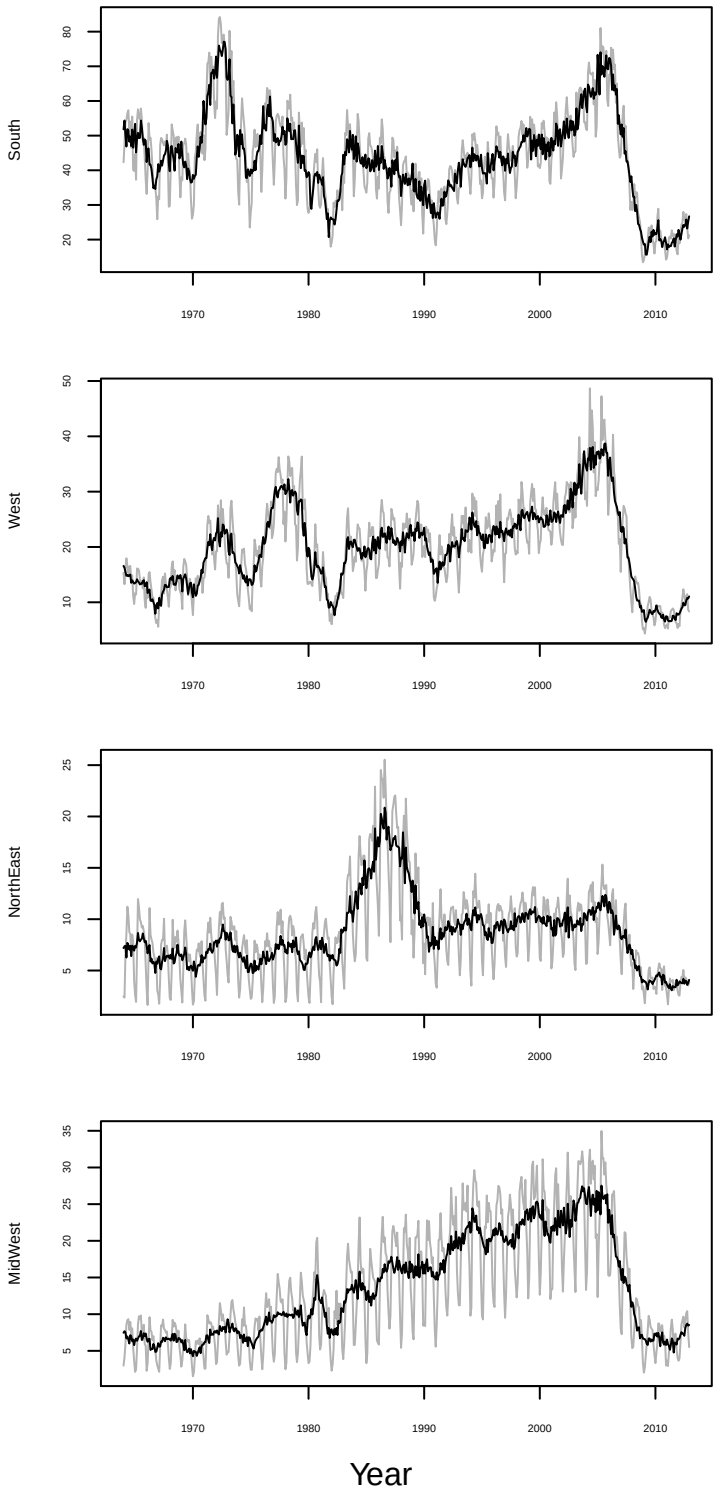

Figure 9: Quadvariate Housing Starts data (1964.1 through 2012.12) in grey, with black MB seasonal adjustments.

zero) and the noise has differencing operator $U(B)$ (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 10. All the in-sample MSEs are provided in Table 3, with results displayed in Figures B.3, B.4, B.5, and B.6 of the Appendix.

	Null Model		Alternative Model		Starts Data	
Series	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 veritable 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

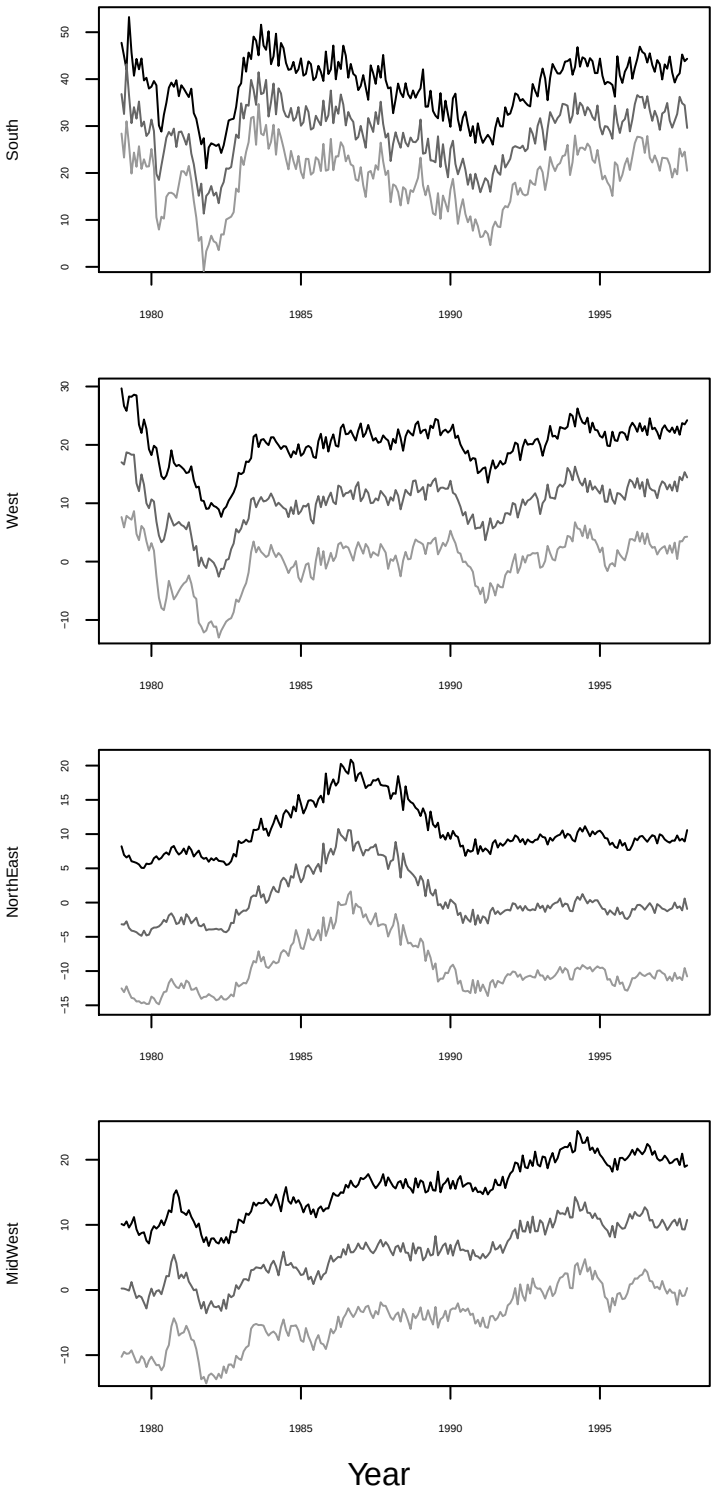

Figure 10: Starts data structural quadvariate seasonal adjustment output (black), with non-stationary MDFA seasonal adjustment (dark grey) and MB concurrent seasonal adjustment (light grey). The seasonal adjustment lines have been vertically staggered for easier visualization.

choice of the target signal, but utilizes a nonparametric spectral estimate in the criterion and is facillitated 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 outperforms the MBA in cases of model mis-specification. Another distinguishing feature of our methoology 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 accomodate 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 at a unit root $\zeta = e^{i\omega}$ can be viewed as the occurence 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 (2018) 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.

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] Baxter, M. and King, R. (1999) Measuring business cycles: approximate bandpass filters for economic time series. *Review of Economics and Statistics* **81**, 575–593.

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

[4] 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.

[5] Brockwell, P. and Davis, R. (1991) *Time Series: Theory and Methods.* New York: Springer.

[6] 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.

[7] Golub, G.H. and Van Loan, C.F. (2012) *Matrix computations (Vol. 3)*. JHU Press.

[8] Harvey, A. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

[9] Lütkepohl, H. (2007) *New Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

[10] 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.

[11] McElroy, T. (2008) Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* **24**(4), 988–1009.

[12] McElroy, T. (2017) Multivariate seasonal adjustment, economic identities, and seasonal taxonomy. *Journal of Business & Economic Statistics* **35**(4), 611–625.

[13] McElroy, T. and McCracken, M. (2017) Multistep ahead forecasting of vector time series. *Econometric Reviews* **36**(5), 495–513.

[14] 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.

[15] Roy, A., McElroy, T., and Linton, P. (2018) Estimation of causal invertible VARMA models. Forthcoming, *Statistica Sinica*.

[16] 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.

[17] Wildi, M. (2008) *Real-Time Signal Extraction: Beyond Maximum Likelihood Principles.* Berlin: Springer.

[18] Wildi, M. and McElroy, T. (2016) Optimal real-time filters for linear prediction problems. *Journal of Time Series Econometrics* **8** (2), 155–192.

[19] Wildi, M. and McElroy, T. (2018) The trilemma between accuracy, timeliness and smoothness in real-time signal extraction. Forthcoming, *International Journal of Forecasting.*