Optimal Real-Time Filters for Linear Prediction Problems

Marc Wildi*and Tucker McElroy†
U.S. Census Bureau

Abstract

The classic model-based paradigm in time series analysis is rooted in the Wold decomposition of the data-generating process into an uncorrelated "white noise" process. By design, this universal decomposition is indifferent to particular features of a specific prediction problem (e.g., forecasting or signal extraction) – or features driven by the priorities of the data-users. A single optimization principle (one-step ahead forecast error minimization) is proposed by this classical paradigm to address a plethora of prediction problems. In contrast, this paper proposes to reconcile prediction problem structures, user priorities, and optimization principles into a general framework whose scope encompasses the classic approach. We introduce the linear prediction problem (LPP), which in turn yields an LPP objective function. Then one can fit models via LPP minimization, or one can directly optimize the linear filter corresponding to the LPP, yielding the Direct Filter Approach. We provide theoretical results and practical algorithms for both applications of the LPP, and discuss the merits and limitations of each. Our empirical illustrations focus on trend estimation (low-pass filtering) and seasonal adjustment in real-time, i.e., constructing filters that depend only on present and past 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

Two applications of great interest in time series analysis are forecasting and signal extraction (cf. Brockwell and Davis (1991, p.8)). A key aspect of forecasting is that no future data can be used, and the same feature holds for concurrent signal extraction problems. When it is required

*IDP, Zurich University of Applied Sciences, Rosenstrasse 8, 8401 Winterthur, Switzerland, marc.wildi@zhaw.ch
†Center for Statistical Research and Methodology, U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C. 20233-9100, tucker.s.mcelroy@census.gov

to compute such projections quickly, without the guidance of cross-validating data, the task is referred to as real-time forecasting/signal extraction. This real-time perspective is in contrast to historical estimators, which take a retrospective view on signal extraction, and may utilize data that is future with respect to the time point under consideration. Considerable applied interest is focused on the real-time analysis of economic time series, as the identification of trends, cycles, and turning points has a tremendous impact on public policy and private investment (Harvey (1989, p.3)). Also, concurrent seasonal adjustment has vast implications on public policy. For a recent discussion of seasonal adjustment in the Great Recession, see Maravall and Pérez (2012). Also see Bell and Hillmer (1984), Findley et al. (1998), Dagum and Luati (2012), and Tiller (2012) for further discussion of seasonal adjustment, and Alexandrov et al. (2012) for a review of trend extraction methods.

It has long been recognized that a trade-off exists between accuracy (or reliability) of real-time methods, and their timeliness (see the discussion in Wildi (2005, 2008)). This tension is best illustrated by the task of finding long-term turning points in economic time series, such as the Industrial Production Index or the Gross Domestic Product. One wishes to accurately find turning points before they occur; the production of forecasted turning points antecedent to their manifestation is highly desirable. Although such estimated turning points are timely, some of them may be spurious, or false, which causes confusion and incorrect decisions. Hence, turning points may be timely but inaccurate. Conversely, it is relatively simple to produce highly accurate real-time turning points that manifest well after the phenomenon has been observed – such estimates are not timely. By expanding the class of real-time filters, and directly minimizing signal extraction mean squared error (as opposed to one-step ahead forecasting error), it is possible to improve performance; this is the main thesis of the paper.

First, in Section 2 we introduce a fairly broad class of linear prediction problems, and discuss classically optimal solutions, where optimality means minimization of the Mean Square Error (MSE) of the real-time estimator. This collection of problems is called the set of Linear Prediction Problems (LPPs). Our results demonstrate that the optimal solution of a LPP depends upon innate characteristics of the time series (through its Wold decomposition), and these might typically be approximated by postulated models. Of course, it is natural to fit these models such that the resulting real-time prediction MSE is minimized, which may very well produce non-classical parameter estimates, i.e., estimates other than Maximum Likelihood Estimates (MLEs) or other efficient estimators, such as Whittle estimates. These alternative methods of fitting are discussed in Section 3, offering a novel generalization of the multi-step ahead forecasting criterion of McElroy and Wildi (2013).

Secondly, we describe in Section 4 a non model-based approach to these prediction problems, which attempts to minimize real-time MSE with respect to some chosen class of concurrent filters – this is called the Direct Filter Approach (DFA), described fully in Wildi (2005, 2008) – with

a resulting methodology that typically differs from classical model-based approaches. Our results connect DFA to the classical approaches, allowing for contrasts to be made. Although the DFA has existed for over a decade, the connections to general time series prediction problems made herein are novel. Moreover, the application of the DFA from a completely model-based orientation is a fresh development.

Section 5 applies these concepts on a few worked examples, demonstrating explicitly the power of accounting for prediction problem structure and user priorities directly in the objective function. User priorities may focus on long-term forecasting, or trend extraction, or seasonal adjustment, or business cycle turning points, for example; these can be encapsulated by an particular LPP, so that the objective function matches the application. We focus on the important U.S. automobile retail sector for an example involving trend estimation in the presence of strong seasonality. We illustrate how the DFA can replicate, or reproduce, classical model-based methods of real-time signal extraction. We then successively change the inputs to the DFA objective function, including the target signal and the spectral estimate. We compare the resulting filter with a widely used model-based design. For a seasonal adjustment exmaple we study U.S. housing starts for the MidWest region. The seasonality of this series has the common feature (among economic data) that its seasonal peaks differ in width and height. We first show how this salient feature of the series can be accounted for, and then compare real-time DFA seasonal adjustment performances with a classical model-based approach. Section 6 concludes, and both code and mathematical proofs are in the Appendix.

In summary, this paper offers three novel contributions: (1) we define and solve LPPs, which generalize simple forecasting and signal extraction problems; (2) we treat model fitting via minimization of LPP MSE, describing the asymptotic properties of parameter estimates and their pseudo-true values; (3) we connect these two previous concepts to the DFA, showing that the DFA is broader, while deriving asymptotic properties of parameter estimates. These three contributions are tied together through two extensive empirical illustrations.

2 MSE Optimal Prediction Problems

We focus in this paper on univariate difference stationary time series, defined below. Throughout, B is the backshift operator and $F = B^{-1}$ is the forward shift operator. The autocovariance function (acf) of a weakly stationary time series with bounded spectral density f (and bounded away from zero, so that long memory and negative memory is excluded) is denoted $\gamma_h(f)$ at lag h, and is defined as the inverse Fourier Transform of the spectrum, i.e.,

$$\gamma_h(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda h} f(\lambda)\, d\lambda.$$

The autocovariance matrix of dimension n is then denoted $\Sigma(f)$, and its jkth entry is $\gamma_{j-k}(f)$. We also use $z = e^{-i\lambda}$ for $\lambda \in [-\pi, \pi]$. In this section we discuss real-time signal extraction and the solution to the Linear Prediction Problem (LPP).

2.1 The Linear Prediction Problem

We begin by defining the class of real-time estimation problems considered in this paper, which are developed through several examples.

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 $\Psi(B)$ acting on a given observed time series $\{X_t\}$ if and only if we can write for all integers t

$$Y_t = \Psi(B)X_t.$$

Throughout this paper we will write the frequency response function (frf) of a linear filter $\Psi(B) = \sum_{j=-\infty}^{\infty} \psi_j B^j$ via $\Psi(z)$, where $z = e^{-i\lambda}$. Thus the frf is a function with domain $\lambda \in [-\pi, \pi]$.

Example 1: One-step Ahead Forecasting. Here the target is X_{t+1}, so that $\Psi(B) = B^{-1}$.

Example 2: Multi-step Ahead Forecasting. Instead we want to project h steps ahead with $h \geq 1$, so $Y_t = X_{t+h} = F^h X_t$, and $\Psi(B) = B^{-h}$.

Example 3: HP Low-pass. The Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997) is a low-pass filter appropriate for producing trends. The output of the filter is our target in this case, and

$$\Psi(z) = \frac{q}{q + (1-z)^2(1-\overline{z})^2}$$

is the frf, where $q > 0$ is the signal-to-noise ratio.

Example 4: HP High-pass. The HP filter is also used to define cycles in the econometric literature, by taking the identity minus the HP low-pass filter. So the target is a cycle and the filter frf is

$$\Psi(z) = \frac{(1-z)^2(1-\overline{z})^2}{q + (1-z)^2(1-\overline{z})^2}.$$

See McElroy (2008) for formulas for the filter coefficients.

Example 5: Naïve Seasonal Adjustment. The removal of seasonal patterns most simply involves an annual summation of past values. Symmetrizing and normalizing to ensure preservation of levels yields the simplistic filter

$$\Psi(B) = s^{-2}U(B)U(F),$$

where s is the number of seasons in the year (e.g., $s = 4$ for quarterly data and $s = 12$ for monthly data) and $U(B) = 1 + B + B^2 + \cdots + B^{s-1}$. As shown in McElroy and Wildi (2010), the seasonal estimation filter $1 - \Psi(B)$ can be expressed as $(1 - B)(1 - F)$ times a symmetric MA filter, which indicates that the seasonal adjustment filter preserves quadratic trends.

Example 6: Henderson Trend. Introduced in actuarial science, the Henderson filter – see Ladiray and Quenneville (2001) for more background – is typically used to produce trends. The coefficients depend on an (odd integer) order q, but all Henderson filters have the form

$$\Psi(B) = 1 - (1 - B)^2(1 - F)^2\Phi^q(B),$$

where Φ^q is a symmetric function of B and F of maximum order $(q - 5)/2$. For example, $\Phi^9(B) = .33 + .17(B + F) + .04(B^2 + F^2)$. Other cases are given in McElroy (2011).

Example 7: X-11 Filters. The trend, seasonal, nonseasonal, and irregular components are defined as the output of an iterative nonlinear procedure in the software program X-11 (Ladiray and Quenneville (2001) describe the procedure). When linearized, the filters can be expressed as symmetric MA filters described in McElroy (2011).

Example 8: Ideal Low-Pass. The concept of the ideal low-pass filter involves a steep cutoff of noise frequencies, described by an indicator function for the frf; see Baxter and King (1999). Thus $\Psi(z) = 1_{[-\mu,\mu]}(\lambda)$ for some cutoff $\mu \in (0, \pi)$ that separates the pass-band from the stop-band. There are infinitely many nonzero coefficients, given by $\psi_j = \sin(j\mu)/(\pi j)$ for $j \neq 0$ and $\psi_0 = \mu/\pi$. (The ideal band-pass filter arises as the difference of two ideal low-pass filters.)

The targets of real-time signal extraction can be forecasts or other features of the process. In general, they represent features of the data process that are of interest to the user. The real-time estimation problem is concerned with projecting the target Y_t onto the available data $X_{t:} = \{X_t, X_{t-1}, \cdots\}$, i.e., the semi-infinite past. We seek a solution that expresses the estimate as a linear combination of the data, or in other words a linear (time-invariant) concurrent filter applied to $\{X_t\}$. We desire that the error in approximating the target with the available data be small.

Although in practice only a finite past is actually available, most real-time filters have coefficients

that decay at geometric rate[1], such that there is little difference between a filter of length 200 and an infinite length filter. That is, if we have at least 200 or so data points, there is generally no loss in simply truncating the semi-infinite real-time filter at the 200th coefficient.

More formally, our estimate of the target Y_t is denoted $\widehat{Y}_t$, and can be expressed via $\widehat{\Psi}(B)X_t$ with $\widehat{\Psi}(B) = \sum_{k\geq 0}\widehat{\psi}_k B^k$, a causal (or concurrent) filter. We refer to this as the "linear time-invariant real-time estimation problem." Note that if the data process were Gaussian, we could view our estimate as the conditional expectation $\widehat{Y}_t = \mathbb{E}[Y_t|X_{t:}]$. Then the coefficients $\{\widehat{\psi}_k\}$ would be selected to minimize the Mean Squared Error (MSE) of the approximation error $Y_t - \widehat{Y}_t$, using the second order properties of the data process $\{X_t\}$ (summarized through its spectral density function $\widetilde{f}$). More generally, our data process might not be Gaussian, but we still seek a causal linear solution because it is convenient for applications; MSE may still be a useful error metric for non-Gaussian processes. See Bell (1984) for further discussion of MSE in signal extraction problems, and McElroy (2010) for alternative measures.

Definition 2 *:* The **Linear Prediction Problem** (LPP) seeks the minimal MSE linear estimate that solves the real-time estimation problem. That is, the LPP involves determining causal $\widehat{\Psi}(B)$ such that the prediction error

$$Y_t - \widehat{Y}_t = \left[\Psi(B) - \widehat{\Psi}(B)\right] X_t$$

has mean zero and minimal MSE.

Example 1: One-step Ahead Forecasting. The LPP in this case refers to determination of optimal forecasts, and

$$Y_t - \widehat{Y}_t = X_{t+1} - \widehat{\Psi}(B)X_t = \left[F - \widehat{\Psi}(B)\right] X_t.$$

Example 2: Multi-step Ahead Forecasting. The LPP is optimal h-step forecasting, and the forecast error is $(F^h - \widehat{\Psi})X_t$.

Example 5: Naïve Seasonal Adjustment. The LPP involves optimal real-time estimation of the simplistic seasonal adjustment. Thus

$$Y_t - \widehat{Y}_t = \left[s^{-2}U(B)U(F) - \widehat{\Psi}(B)\right] X_t.$$

[1]Filters that are derived from models inherit the properties of these models, so that short memory models induce filters with coefficients of geometric decay (like the autocovariances), whereas long memory models can produce filter coefficients with hyperbolic decay – e.g., see Holan and McElroy (2012). We focus on processes with bounded spectrum, and short memory models, so that filter coefficients decay rapidly. Whereas the methodology of this paper does not preclude long memory, the mathematical results (and neither the empirical results) have not been established by us for a long memory framework.

We note here that although our forecasting LPPs are conventional, signal extraction is often (see Bell and Hillmer (1984)) formulated in terms of unobserved stochastic processes, where the target is not expressible as a linear filter of the data. The perspective on signal extraction in this paper is different, and is equivalent to revision minimization (of the semi-infinite to the bi-infinite filters) in the classical paradigm.

2.2 Solution to the Linear Predction Problem

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. We suppose that there exists a differencing polynomial $\delta(B)$ such that $W_t = \delta(B)X_t$ is a covariance stationary time series. Here δ is a degree d polynomial with all its roots on the unit circle of the complex plane. All purely nondeterministic stationary (mean zero) processes have a Wold decomposition $W_t = \Pi(B)\epsilon_t$, where $\{\epsilon_t\}$ is white noise (uncorrelated serially, but possibly dependent over time) of variance σ^2 and $\Pi(B) = \sum_{j=-\infty}^{\infty} \pi_j B^j$ (Brockwell and Davis, 1991). When $\pi_j = 0$ for all $j < 0$, the process is called causal. For any Laurent series $\Upsilon(B) = \sum_j \upsilon_j B^j$, let the notation $[\Upsilon]_r^s(B)$ denote $\sum_{j=r}^{s} \upsilon_j B^j$.

We begin our treatment with some preliminary results from Bell (1984) on nonstationary stochastic processes. Let $\delta(z) = 1 - \sum_{j=1}^{d} \delta_j z^j$, and its reciprocal power series is $\xi(z) = 1/\delta(z) = \sum_{j\geq 0} \xi_j z^j$. One can recursively solve for the $\{\xi_j\}$ via $\xi_0 = 1$ and $\xi_j = \sum_{k=1}^{\min(d,j)} \delta_k \xi_{j-k}$ for $j \geq 1$. Moreover, certain time-dependent coefficient functions $A_{j,t}$ lying in the null space of $\delta(B)$ are defined via

$$A_{j,t} = \xi_{t-j} - \sum_{k=1}^{d-j} \delta_k \xi_{t-j-k}$$

for $j = 1, 2, \cdots, d$ and $t \geq 1$. Then the process $\{X_t\}$ can be represented via

$$X_t = \sum_{j=1}^{d} A_{j,d+t} X_{j-d} + \sum_{j=0}^{t-1} \xi_j W_{t-j}. \tag{1}$$

It also follows from results in Bell (1984) that

$$1 - \sum_{j=1}^{d} A_{j,d+t} B^{d+t-j} = \sum_{k=0}^{t-1} \xi_k B^k \, \delta(B), \tag{2}$$

which is an algebraic identity. Assuming the spectral representation for $\{W_t\}$ exists (see Brockwell and Davis (1991) for additional details), namely $W_t = \int_{-\pi}^{\pi} e^{i\lambda t} \, d\mathbb{Z}(\lambda)$ for an orthogonal increments process $\mathbb{Z}(\lambda)$, we obtain the following spectral representation for $\{X_t\}$:

$$X_t = \sum_{j=1}^{d} A_{j,d+t} X_{j-d} + \int_{-\pi}^{\pi} \frac{e^{i\lambda t} - \sum_{j=1}^{d} A_{j,d+t} e^{-i\lambda(d-j)}}{\delta(z)} \, d\mathbb{Z}(\lambda). \tag{3}$$

This expresses the dynamics of the process in terms of a predictable portion – determined by the functions $A_{j,d+t}$ and the variables $X_{1-d}, \cdots, X_{-1}, X_0$ – and a non-predictable portion involving a

time-varying filter of the $\{W_t\}$ series. Then a target signal, given the application of a linear filter $\Psi(B)$, takes the form

$$Y_t = \Psi(B)X_t = \sum_{j=1}^{d} \Psi A_{j,d+t} X_{j-d} + \int_{-\pi}^{\pi} \frac{e^{i\lambda t}\Psi(z) - \sum_{j=1}^{d} \Psi A_{j,d+t} e^{-i\lambda(d-j)}}{\delta(z)} \, d\mathbb{Z}(\lambda),$$

where $\Psi A_{j,d+t} = \sum_k \psi_k A_{j,d+t-k}$ describes the action of the filter on the coefficient functions. The LPP requires that the filter error process $Y_t - \widehat{Y}_t$ be mean zero, in addition to having minimal variance. The mean zero condition is automatic for stationary data (when $d = 0$, the empty sums collapse and the mathematics is much simpler).

We first describe a broad set of conditions that *any* real-time signal extraction filter must satisfy to even qualify as a solution to the LPP. Essentially, the filter error $\Psi(B) - \widehat{\Psi}(B)$ must be divisible by $\delta(B)$, as shown below. First, if we write down $Y_t - \widehat{Y}_t = [\Psi(B) - \widehat{\Psi}(B)]X_t$, we have a potentially nonzero mean arising from terms of the form $\Psi A_{j,d+t} - \widehat{\Psi} A_{j,d+t}$. We require that these quantities be identically zero, i.e., the real-time filter exactly replicates the behavior of $\Psi(B)$ on underlying predictable components of the data process. Since the functions $\{A_{j,\cdot}\}_{j=1}^{d}$ constitute a basis for the Null Space of the operator $\delta(B)$, it suffices to determine some linear filter $\tau(B)$ such that $\Psi(B) - \widehat{\Psi}(B) = \tau(B)\delta(B)$.

Suppose that $\delta(B) = \Pi_\ell (1 - B\zeta_\ell^{-1})^{r_\ell}$, which factors the differencing operator in terms of roots ζ_ℓ occurring with multiplicity r_ℓ. Then we impose

$$\widehat{\Psi}^{(r_\ell - 1)}(\zeta_\ell) = \Psi^{(r_\ell - 1)}(\zeta_\ell) \tag{4}$$

for each ℓ. This notation says that the derivative of order $r_\ell - 1$ of the Laurent series, evaluated at the corresponding root ζ_ℓ with that multiplicity $r_\ell - 1$, is the same for both Ψ and $\widehat{\Psi}$ by fiat. Let $\Delta(B) = \Psi(B) - \widehat{\Psi}(B)$, and observe that (4) ensures that $\Delta^{(r_\ell - 1)}(\zeta_\ell) = 0$ for all ℓ. By the unique factorization of polynomials over the complex plane, it follows that (4) guarantees that $\Delta(B)$ is divisible by $\delta(B)$; we denote the quotient by $\tau(B)$. As a result, the real-time error process is

$$Y_t - \widehat{Y}_t = \Delta(B)X_t = \tau(B)W_t$$

when $\widehat{\Psi}$ is selected to satisfy (4). This ensures that the error process is mean zero (and covariance stationary). With these preliminaries, we can state the solution to the LPP.

Proposition 1 *Suppose that $\{X_t\}$ is nonstationary with representation (3), and that $\{W_t\}$ is causal, expressed as $W_t = \Pi(B)\epsilon_t$. Moreover, assume that the initial values $X_0, \cdots, X_{1-d}$ are uncorrelated with the innovations $\{\epsilon_t\}$. Then the solution to the LPP posed by a given $\Psi(B)$ is given by*

$$\widehat{\Psi}(B) = \sum_{h \geq 0} \psi_h B^h + \sum_{h<0} \psi_h \left(\sum_{j=1}^{d} A_{j,d-h} B^{d-j} + \sum_{k=1}^{-h} \xi_{-h-k} [\Pi]_k^{\infty}(B) F^k \delta(B) \Pi^{-1}(B) \right). \tag{5}$$

Moreover, the minimal MSE is given by

$$\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\left|\sum_{h>0}\psi_{-h}z^{-h}[\Pi/\delta]_0^{h-1}(z)\right|^2 d\lambda. \tag{6}$$

Remark 1 Implicit in the proof is the fact that the error filter $\Psi(B)-\widehat{\Psi}(B)$ is divisible by $\delta(B)$. In general, the causal moving average $\Pi(B)$ is unknown to us, and we may attempt various guesses at its structure, typically via utilizing fitted models. The formula (6) gives us a lower bound on the MSE when we use sub-optimal proxies for $\widehat{\Psi}(B)$.

As indicated by Remark 1, the result of Proposition 1 is only useful if we know $\Pi(B)$, or have some decent approximation. A classical approach would be to formulate a model for $\Pi(B)$, compute the LPP MSE as a function of model parameters, and minimize this function to determine the best possible $\Pi(B)$ for that model class. Or we might determine model parameters some other way (e.g., through MLEs) and plug into the formula. We pursue these ideas further in the next Section.

3 Model Fitting via LPP MSE Minimization

In this section we use the variance of the LPP to fit models, making connections to the Whittle likelihood and Kullback-Leibler discrepancy; see Taniguchi and Kakizawa (2000). Let us suppose that a model is postulated for the data process, which can be visualized by considering a particular class of $\Pi_\omega(B)$ parameterized by a vector $\omega\in\Omega$, a model parameter manifold. (Note that the innovation variance σ^2 is not considered part of the parameter vector ω, as we focus on separable models, i.e., the innovation variance is separately parametrized.) We presume that the unit roots – encapsulated in δ – have been correctly identified. The model spectral density (for the differenced data process) is then $|\Pi_\omega(z)|^2\sigma^2$, denoted by $f_\omega(\lambda)$. The "innovation-free" spectrum is defined as $\overline{f}_\omega(\lambda)=|\Pi_\omega(z)|^2=f_\omega(\lambda)/\sigma^2$. The modeler hopes that f_ω forms a suitable approximation to $\widetilde{f}$ (the true spectrum of the differenced data process) once the parameter ω is appropriately fitted. In practice, this involves finding $\widetilde{\omega}\in\Omega$ such that $f_{\widetilde{\omega}}$ and $\widetilde{f}$ are close according to some distance metric. The empirical version then chooses $\widehat{\omega}\in\Omega$ such that $f_{\widehat{\omega}}$ and I (the periodogram) are close according to the same metric. The periodogram is computed from a sample of size n from the (stationary differenced) process, namely $W=(W_1,W_2,\cdots,W_n)'$, and is defined by $I(\lambda)=n^{-1}|\sum_{t=1}^n W_t z^t|^2$. Observe that $n^{-1}W'\Sigma(g)W=\gamma_0(gI)$ for any function g defined on the domain $[-\pi,\pi]$; see McElroy and Holan (2009) for derivations.

For further exposition of this basic approach, see Taniguchi and Kakizawa (2000) and McElroy and Wildi (2013). The latter paper considers the multi-step ahead LPP (Example 2 above). In the case of the one-step ahead LPP, the MSE of the LPP error corresponds to the Whittle likelihood (up to a term involving the log innovation variance) and is related to Kullback-Leibler (KL) discrepancy (Dahlhaus and Wefelmeyer, 1996). We provide a general, and novel, treatment of this topic below.

When using a potentially misspecified model to solve a LPP, the real MSE is the variance of

$$\eta_t = \sum_{h>0} \psi_{-h} F^h [\Pi_\omega/\delta]_0^{h-1}(B) \Pi_\omega^{-1}(B) W_t$$

so long as the unit roots are correctly identified (see the proof of Proposition 1). Since $\Pi_\omega(B)$ is now potentially misspecified, we cannot conclude that $\Pi_\omega^{-1}(B)W_t = \epsilon_t$ as in the proof of Proposition 1. Elementary calculations then yield

$$Var(\eta_t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{h>0} \psi_{-h} z^{-h} [\Pi_\omega/\delta]_0^{h-1}(z) \right|^2 \frac{\widetilde{f}(\lambda)}{\overline{f}_\omega(\lambda)} \, d\lambda. \tag{7}$$

Note that (7) then becomes a function of the model parameter ω, as well as the data spectrum $\widetilde{f}$. Clearly, one comes as close as possible to the optimal MSE target by finding ω to minimize this criterion. Let $\omega(\widetilde{f})$ denote a minimizer of (7), which of course depends on the true spectrum $\widetilde{f}$. Then using the LPP filter $\widehat{\Psi}(B)$ corresponding to this particular $\omega(\widetilde{f})$ provides the best possible concurrent approximation to $\Psi(B)$ within the given model. It will be convenient to generalize (7) to a function $J_\Psi(\omega, g)$:

$$J_\Psi(\omega, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{h>0} \psi_{-h} z^{-h} [\Pi_\omega/\delta]_0^{h-1}(z) \right|^2 \frac{g(\lambda)}{\overline{f}_\omega(\lambda)} \, d\lambda.$$

Here g is a generic real-valued non-negative function with domain $[-\pi, \pi]$. This J_Ψ provides a distance measure between the functions g and $\overline{f}_\omega$ as a function of ω (through Π_ω), depending on the given Ψ. Its minimizer (with respect to $\omega \in \Omega$) is denoted $\omega(g)$. When $g = \widetilde{f}$, then $\omega(\widetilde{f})$ provides the lowest possible LPP MSE for that particular data process. But if $g = I$, then $\omega(I)$ provides an empirical estimate of the $\omega(\widetilde{f})$, as shown below. We refer to $\omega(\widetilde{f})$ as a pseudo-true value (PTV), in analogy with the terminology used for the Whittle likelihood and the KL discrepancy. (See Cox (1961, 1962) for further background on PTVs.)

Now when the model is correctly specified, there must exist some "true" parameter $\widetilde{\omega} \in \Omega$ such that $\widetilde{f} \propto f_{\widetilde{\omega}}$ because $\widetilde{f}$, once divided by its innovation variance $\widetilde{\sigma}^2 = \exp\{(2\pi)^{-1} \int_{-\pi}^{\pi} \log \widetilde{f}(\lambda) \, d\lambda\}$, is in $\{f_\omega : \omega \in \Omega\}$. (There may be multiple such true parameters $\widetilde{\omega}$ if there is an identifiability problem.) Then the identity $\widetilde{f} = \overline{f}_{\widetilde{\omega}} \, \widetilde{\sigma}^2$ holds, and it follows that

$$J_\Psi(\overline{\omega}, \widetilde{f}) = \frac{\widetilde{\sigma}^2}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{h>0} \psi_{-h} z^{-h} [\Pi_{\widetilde{\omega}}/\delta]_0^{h-1}(z) \right|^2 d\lambda.$$

By Remark 1, this quantity achieves the minimal MSE lower bound of the LPP. But because by definition $J_\Psi(\omega(\widetilde{f}), \widetilde{f}) \leq J_\Psi(\omega, \widetilde{f})$ for all $\omega \in \Omega$, we must have $J_\Psi(\omega(\widetilde{f}), \widetilde{f}) = J_\Psi(\widetilde{\omega}, \widetilde{f})$. Then if the function $J_\Psi(\cdot, \widetilde{f})$ has a unique minimizer, we conclude that $\omega(\widetilde{f}) = \widetilde{\omega}$, i.e., the minimizer of the LPP criterion is identical with the true parameter.

More generally, the model may be incorrectly specified (i.e., $\widetilde{f} \notin \{f_\omega : \omega \in \Omega\}$), and $\omega(\widetilde{f})$ no longer equals the "true" parameter – and in fact, the concept of "true" parameter becomes an absurd concept. But $\omega(\widetilde{f})$ comes as close as possible to truth, according to the given metric, and this justifies the nomenclature of "pseudo-true value" (PTV). We now formulate a general result about inference for model parameters ω based on the LPP MSE.

We must assume that our PTVs are not on the boundary of the parameter set, because the limit theory is non-standard in this case (cf. Self and Liang (1987)). If the PTV is unique, the Hessian of the criterion function should be positive definite at that value, and hence invertible. The so-called Hosoya-Taniguchi (HT) conditions (Hosoya and Taniguchi (1982) and Taniguchi and Kakizawa (2000)) impose sufficient regularity on the process $\{W_t\}$ to ensure a central limit theorem; these conditions require that the process is a causal filter of a higher-order martingale difference. Finally, we suppose that the fourth order cumulant function of the process is identically zero, which says that in terms of second and fourth order structure the process looks Gaussian. This condition is not strictly necessary, but facilitates a simple expression for the asymptotic variance of the parameter estimates. Let the Hessian of $J_\Psi(\omega, \widetilde{f})$ be denoted $H(\omega) = \nabla_\omega \nabla'_\omega J_\Psi(\omega, \widetilde{f})$.

Theorem 1 *Suppose that $\omega(\widetilde{f})$ exists uniquely in the interior of Ω and that $H(\omega(\widetilde{f}))$ is invertible. Suppose that the process $\{W_t\}$ has finite fourth moments, conditions (HT1)-(HT6) of Taniguchi and Kakizawa (2000, pp.55-56) hold, and that the fourth order cumulant function of $\{W_t\}$ is zero. Then as $n \to \infty$*

$$\begin{aligned} \omega(I) &\xrightarrow{P} \omega(\widetilde{f}) \\ \sqrt{n}\left(\omega(I) - \omega(\widetilde{f})\right) &\xRightarrow{\mathcal{L}} \mathcal{N}\left(0, H^{-1}(\omega(\widetilde{f}))V(\omega(\widetilde{f}))H^{-1}(\omega(\widetilde{f}))\right). \end{aligned}$$

Here V is a matrix given by

$$\begin{aligned} V(\omega) &= \frac{2}{2\pi}\int_{-\pi}^{\pi} \nabla r_\omega(\lambda) \cdot \nabla' r_\omega(\lambda)\, d\lambda \\ r_\omega(\lambda) &= \frac{\left|\sum_{h>0} \psi_{-h} z^{-h} [\Pi_\omega/\delta]_0^{h-1}(z)\right|^2}{\overline{f}_\omega(\lambda)}. \end{aligned}$$

This result illustrates a form of consistency of the estimates, as well as illustrating when and how efficiency fails. The asymptotic normality result can be used to build models, since we can compute the asymptotic variances and use these quantities to test whether estimated parameters are significantly different from zero. Of course, this depends on our ability to compute V and H.

Example 1: One-step Ahead Forecasting. With g equal either to the periodogram I or the spectrum $\widetilde{f}$, $J_\Psi(\omega, g) = (2\pi)^{-1}\int_{-\pi}^{\pi} g(\lambda)/\overline{f}_\omega(\lambda)\, d\lambda$ is the MSE for one-step ahead forecasting (the differencing operator δ plays no role). Modulo the contribution of the innovation variance, this is the Whittle likelihood, and is also the KL discrepancy between g and $\overline{f}_\omega$ (Taniguchi and

Kakizawa, 2000). The estimate ω_I is called the quasi-maximum likelihood estimate (QMLE), or Whittle estimate; in the case of an AR(p) model, it is the solution to the empirical Yule-Walker equations (see McElroy and Findley (2010) for more discussion). More explicitly,

$$\overline{f}_{\omega}(\lambda) = |1 - \omega_1 z - \cdots - \omega_p z^p|^{-2}$$

and $\omega = [\omega_1, \omega_2, \cdots, \omega_p]'$. The MSE function is then explictly written

$$J_\Psi(\omega, g) = \gamma_0(g) - 2[\gamma_1(g), \gamma_2(g), \cdots, \gamma_p(g)]\, \omega + \omega' \Sigma(g) \omega,$$

so that the optimum is $\omega_g = \Sigma^{-1}(g)[\gamma_1(g), \gamma_2(g), \cdots, \gamma_p(g)]'$. This is the familiar solution to the Yule-Walker equations.

Example 2: Multi-step Ahead Forecasting. Now the LPP criterion is

$$J_\Psi(\omega, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|[\Pi_\omega/\delta]_0^{h-1}(z)|^2}{|\Pi_\omega(z)|^2}\, g(\lambda)\, d\lambda,$$

which is the h-step ahead prediction MSE discussed in McElroy and Findley (2010) and McElroy and Wildi (2013). The latter paper provides an explicit expression in the case of a fitted ARIMA(1,1,0) model. In this case the model is written $(1-B)(1-\omega B)X_t$ is a white noise process. Then

$$J_\Psi(\omega, g) = \gamma_0(g)\, [h + \zeta^2(\omega)] + 2 \sum_{k=1}^{h} \gamma_k(g)\, [h - k + \zeta(\omega)]$$
$$\zeta(\omega) = -\omega\, \frac{1 - \omega^h}{1 - \omega},$$

where $g = I$ or $\widetilde{f}$ as the case may be. An optimizer $\omega(g)$ is any root of $\omega + \omega^2 + \cdots + \omega^h = \sum_{k=1}^h \rho_k(g)$.

Example 8: Ideal Low-Pass. The best model-based real-time filter approximation to the low-pass filter has MSE

$$J_\Psi(\omega, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{|\sum_{h \geq 1} \frac{\sin(h\mu)}{\pi h} z^{-h} [\Pi_\omega/\delta]_0^{h-1}(z)|^2}{|\Pi_\omega(z)|^2}\, g(\lambda)\, d\lambda.$$

These examples show the connection between the LPP objective functions for model-fitting and the classical objective functions, such as the Whittle likelihood. The next section dispenses with models, and attempts to directly obtain $\widehat{\Psi}$ from a suite of filters.

4 The Direct Filter Approach

In this section we provide a more generic solution to the LPP by generalizing the class of concurrent filters. One view of Proposition 1 is that it provides a certain class of concurrent filters, namely

those that arise from specified models. But there is no requirement to restrict to such classes of filters – it may be possible to improve performance by utilizing other classes of concurrent filters. Perhaps we do not believe that $\{W_t\}$ has a causal representation, or perhaps we entertain little hope of obtaining a viable model for $\Pi(B)$. Instead, we can choose a class of concurrent filters for $\widehat{\Psi}(B)$ in the LPP and optimize the resulting MSE. We proceed to develop this novel approach to the problem.

Suppose that a class of concurrent filters $\mathcal{G}$ is considered, and is parametrized by a filter parameter $\theta \in \Theta$, a parameter manifold. So $\mathcal{G} = \{\widehat{\Psi}_\theta : \theta \in \Theta\}$. Whereas the model-based approach to the LPP involves minimizing $J_\Psi(\omega, g)$ – here g equals I for the empirical problem, and g equals $\widetilde{f}$ for the theoretical problem – instead the filter-based approach to the LPP involves minimizing the MSE of $Y_t - \widehat{Y}_t$ with respect to filters in $\mathcal{G}$; in order for this filter error to have mean zero, we require that these filters satisfy (4). Obviously, this requires firstly a knowledge of the unit roots present in the data, and secondly the ability to compute the appropriate derivatives of $\Psi(B)$. Then the resulting Direct Filter Approach (DFA) MSE can be written as $G_\Psi(\theta, \widetilde{f})$, where G_Ψ is defined as

$$G_\Psi(\theta, g) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\left|\Psi(z) - \widehat{\Psi}(z)\right|^2}{\left|\delta(z)\right|^2} g(\lambda) d\lambda. \tag{8}$$

The integrand in this expression is well-defined due to the imposed conditions (4). Note that if let $g = \widetilde{f}$ and we associate the denominator $|\delta(z)|^2$ with the spectral density $\widetilde{f}$, the DFA MSE involves the magnitude squared of the filter error $\Psi(z) - \widehat{\Psi}(z)$ multiplied by the pseudo-spectral density $\widetilde{f}(\lambda)|\delta(z)|^{-2}$.

Now the minimizer $\theta(g)$ of the DFA MSE provides the optimal concurrent filter $\widehat{\Psi}_{\theta(g)}$ within the class $\mathcal{G}$ (subject to (4)). In the classical case (previous section) we optimize over model parameters, while in the DFA case we optimize over filter parameters. Moreover, we can view the first case as included in the second case, where the class of filters $\mathcal{G}$ considered consists of those of the form (5), where we identify θ directly with ω.

However, we are free to take less restrictive classes for $\mathcal{G}$ in the hope of obtaining a richer class of filters, and thereby to diminish the LPP MSE. For example, for stationary time series $\mathcal{G}$ could consist of all MA filters of a certain order, with θ denoting the coefficients (we will refer to this as $\mathcal{M}_q$, where the MA filters have order q). Alternatively, $\mathcal{G}$ might consist of all ARMA filters of a particular AR and MA order, or might consist of all Zero-Pole Combination (ZPC) filters of a given specification (Wildi, 2008). The DFA of Wildi (2008) approached the minimization of $G_\Psi(\theta, g)$ over a class $\mathcal{G}$ of appropriately restricted ZPC filters. But here we use the term DFA more broadly to refer to the minimization of $G_\Psi(\theta, g)$ with respect to any desired filter class $\mathcal{G}$.

In the case of one-step ahead forecasting of stationary time series and $\mathcal{G} = \mathcal{M}_q$, the DFA is identical with the model-based LPP solution utilizing an AR($q+1$), as we demonstrate next. Recall that $\Psi(B) = B^{-1}$. For any $\widehat{\Psi}_\theta \in \mathcal{M}_q$, write the parametrization as $\widehat{\Psi}_\theta(B) = \theta_0 + \theta_1 B + \cdots + \theta_q B^q$.

Then the DFA MSE is

$$\begin{aligned} G_\Psi(\theta, g) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |z^{-1} - \widehat{\Psi}(z)|^2 g(\lambda)\, d\lambda \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} |1 - \theta_0 z - \theta_1 z^2 - \cdots - \theta_q z^{q+1}|^2 g(\lambda)\, d\lambda \\ &= \gamma_0(g) - 2\left[\gamma_1(g), \cdots, \gamma_{q+1}(g)\right]\theta + \theta'\Sigma(g)\theta, \end{aligned}$$

where $\theta = [\theta_0, \theta_1, \cdots, \theta_q]'$. The optimizer is $\theta(g) = \Sigma^{-1}(g)\left[\gamma_1(g), \cdots, \gamma_{q+1}(g)\right]'$, which is identical to the solution $\omega(g)$ of the AR(q+1) case of Example 1. Hence the filter-based and model-based criteria are mathematically identical, and the optima are the same. This is a very special case, which works essentially because $\Psi(z)$ has unit magnitude at all frequencies (a similar result holds for the multi-step forecasting LPP, although a constrained AR$(h+q)$ is the result).

Analogously to the use of LPP to fit models, we can also develop an inference theory for the DFA. The following concepts are similarly treated in Wildi (2008). Given a filter class $\mathcal{G}$, the best possible concurrent filter is some $\widehat{\Psi}_{\theta(\widetilde{f})}$ where $\theta(\widetilde{f})$ minimizes $G_\Psi(\theta, \widetilde{f})$. We will call this $\theta_{\widetilde{f}}$ the PTV for the filter parameter, in analogy with the terminology for model parameters, although there is no real "true" filter parameter conceptually in place. Likewise, a natural empirical estimate of this optimal filter parameter is $\theta(I)$, obtained by minimizing $G_\Psi(\theta, I)$. Then we have the following result.

Theorem 2 *Suppose that $\theta(\widetilde{f})$ exists uniquely in the interior of Θ and that the Hessian of G_Ψ, denoted $H(\theta) = \nabla\nabla' G_\Psi(\theta, \widetilde{f})$, is invertible. Suppose that the process $\{W_t\}$ has finite fourth moments, conditions (HT1)-(HT6) of Taniguchi and Kakizawa (2000, pp.55-56) hold, and that the fourth order cumulant function of $\{W_t\}$ is zero. Then as $n \to \infty$*

$$\begin{aligned} \theta(I) &\xrightarrow{P} \theta(\widetilde{f}) \\ \sqrt{n}\left(\theta(I) - \theta(\widetilde{f})\right) &\xRightarrow{\mathcal{L}} \mathcal{N}\left(0, H^{-1}(\theta(\widetilde{f}))V(\theta(\widetilde{f}))H^{-1}(\theta(\widetilde{f}))\right). \end{aligned}$$

Here V is a matrix given by

$$\begin{aligned} V(\theta) &= \frac{2}{2\pi}\int_{-\pi}^{\pi} \nabla r_\theta(\lambda) \cdot \nabla' r_\theta(\lambda)\, d\lambda \\ r_\theta(\lambda) &= |\Psi(z) - \widehat{\Psi}_\theta(z)|^2 |\delta(z)|^{-2}. \end{aligned}$$

The approach utilizing θ_I to determine our empirical filter is motivated by the fact that we can re-express $G_\Psi(\theta, I)$ as follows:

$$G_\Psi(\theta, I) = n^{-1} W' \Sigma \left(|\Psi(z) - \widehat{\Psi}_\theta(z)|^2 |\delta(z)|^{-2}\right) W.$$

Up to stochastic errors tending to zero in probability – the analysis follows from results in Brockwell and Davis (1991), as discussed in Wildi (2008) – this is equal to the average sum of squares of the real-time filter errors $Y_t - \widehat{Y}_t$. Minimizing such a quantity with respect to θ is very natural.

However, it may be felt that a model-based estimate of the spectral density is superior to the periodogram in terms of capturing frequency domain characteristics. Suppose we have a model-based (innovation-free) spectral density $\overline{f}_\xi$ arising from a separable model parametrized by $\xi \in \Xi$. This model class may have no relationship whatsoever to the models considered for generating a class of concurrent filters, as in the previous subsection, and so we use ξ rather than ω to distinguish. We suppose that the spectrum has already been fitted, perhaps via MLE or QMLE, so that $\overline{f}_{\widehat{\xi}}$ is available (where $\widehat{\xi}$ is the parameter estimate). Then our empirical DFA criterion is $G_\Psi(\theta, \overline{f}_{\widehat{\xi}})$, and our corresponding DFA empirical optimum is denoted $\theta(\overline{f}_{\widehat{\xi}})$.

Now the estimate $\widehat{\xi}$ will converge, under suitable conditions and assumptions, to the pseudo-true value $\widetilde{\xi}$ pertaining to the particular model class and fitting function (e.g., the Whittle likelihood). By continuity, the DFA estimate $\theta(\overline{f}_{\widehat{\xi}})$ should converge to $\theta(\overline{f}_{\widetilde{\xi}})$, though naturally we desire convergence to $\theta_{\widetilde{f}}$. These two quantities differ when the model is misspecified, and there will be under-performance of the DFA. Nevertheless, the actual bias that arises may be low, arguing in favor of using a model-based spectrum instead of the periodogram. We formalize these ideas in the following theorem.

Theorem 3 *Suppose that $\widehat{\xi}$ is a parameter estimate such that $\sqrt{n}(\widehat{\xi} - \widetilde{\xi}) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}(0, W(\widetilde{\xi}))$ for some positive definite matrix W. Also suppose that the function $\xi \mapsto \theta(\overline{f}_\xi)$ is continuously differentiable, with value denoted by $L(\xi)$. Then*

$$\theta(\overline{f}_{\widehat{\xi}}) \stackrel{P}{\longrightarrow} \theta(\overline{f}_{\widetilde{\xi}})$$
$$\sqrt{n}\left(\theta(\overline{f}_{\widehat{\xi}}) - \theta(\overline{f}_{\widetilde{\xi}})\right) \stackrel{\mathcal{L}}{\Longrightarrow} \mathcal{N}\left(0, \nabla' L(\widetilde{\xi}) W(\widetilde{\xi}) \nabla L(\widetilde{\xi})\right)$$

as $n \to \infty$. Here ∇L is a matrix with $jkth$ entry given by $\partial \xi_j L_k(\xi)$, the jth derivative of the kth component function in L.

Remark 2 We are really interested in $\theta(\overline{f}_{\widehat{\xi}}) - \theta(\widetilde{f})$, and so using the above central limit theorem, we have an asymptotic bias of $\theta(\overline{f}_{\widetilde{\xi}}) - \theta(\widetilde{f})$. This is clearly zero when the model f_ξ is correctly specified, and may even be zero when the model is misspecified – this is contingent on how the optimum $\theta(g)$ depends upon a given g.

We call the resulting (linear) prediction function $\widehat{\Psi}_{\widehat{\theta}}$ – where $\widehat{\theta}$ is either $\theta(I)$ or $\theta(f_{\widehat{\xi}})$, depending on the approach of either Theorem 2 or 3, respectively – a Linear Prediction Filter (LPF). We distinguish the cases, calling $\widehat{\Psi}_{\theta(I)}$ the empirical LPF, whereas $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}$ is the model-based LPF. We next provide some discussion of these two types of DFA.

In the case of the empirical LPF, we have a broad form of consistency, in that the limiting MSE has the formula $G_\Psi(\theta(\widetilde{f}), \widetilde{f})$, which is as close as possible to the theoretical lower bound (6) given the class of filters $\mathcal{G}$. Recall that the model-based approach of the previous section provides $J_\Psi(\omega(\widetilde{f}), \widetilde{f})$ as the minimal MSE, which ultimately is a special case of the DFA where a particular

(model-based) class of filters $\mathcal{G}$ is utilized. In the case that the true optimal filter given by (5) is included in $\mathcal{G}$, then $\theta(\widetilde{f})$ is such that $\widehat{\Psi}_{\theta(\widetilde{f})}$ is identically equal to the optimal filter, and our MSE is minimal. More generally, we cannot know whether the optimal real-time filter is in $\mathcal{G}$, but we can hope that its distance from $\mathcal{G}$ (with the metric determined with respect to G_Ψ) is small if $\mathcal{G}$ is sufficiently rich.

On the other hand, the model-based LPF can go wrong in an additional way: even if $\mathcal{G}$ includes the optimal real-time filter, so that in a sense $\theta(\widetilde{f})$ corresponds to "truth", we may yet have $\theta(\overline{f}_{\widetilde{\xi}}) \neq \theta(\widetilde{f})$ if the model $\overline{f}_\xi$ is mis-specified. This is a drawback to this method. Moreover, identification of the mapping L seems to be analytically intractable, so that the uncertainty in $\theta(\overline{f}_{\widehat{\xi}})$ cannot be measured in many practical applications. However, the model-based LPF holds a certain appeal: if Ψ corresponds to a LPP derived from the fitted model $\overline{f}_{\widehat{\xi}}$, then $\theta(\overline{f}_{\widehat{\xi}})$ necessarily corresponds to the optimal filter in equation (5) – under the assumption that $\overline{f}_\xi$ is correctly specified – with $\widehat{\xi}$ plugged in to all appropriate quantities. So if one believes the specified model $\overline{f}_\xi$ is correct, then the model-based DFA at once produces the optimal filter (if we pretend that $\widehat{\xi}$ is non-random for the purpose of computing MSEs). Moreover, as sample size increases our DFA filter will converge automatically to $\Psi_{\theta(\overline{f}_{\widetilde{\xi}})} = \Psi_{\theta(\widetilde{f})}$. In this sense, the DFA can be made to reproduce the same results as equation (5) in Proposition 1, merely by setting $g \propto \overline{f}_\xi$. We therefore say that this version of the DFA, i.e., the model-based LPF, can replicate purely model-based (classical) results. Beyond demonstrating the scope of the DFA, the principal appeal of replicating model-based approaches resides in the possibility of modifying optimization criteria to particular user priorities.

5 Applications: Real-Time Trend Extraction and Seasonal Adjustment

5.1 Real-Time Trend Extraction

For an illustration of the methodology proposed in Section 4, we consider an application to the "Auto and Other Motor Vehicles" series[2] ("auto-sales" for short). As claimed by some economists, this time series is a "key cyclical indicator and early barometer for the economic effects of higher oil prices"; see Econbrowser 2012 at `http://www.econbrowser.com/archives/2012/02/economic_condit.html` Accordingly, we organize the empirical analysis with the goal of extracting relevant economic signals, possibly anticipating economic downturns in real-time. We emphasize that our intention here is to illustrate the flexible features of the LPP at the analyst's disposal; we intentionally omit discussions of the business cycle and the design of indicators, as this would take us too far away from the main topic of the article. Our procedure starts with a replication of a pure model-

[2] Total sales, based on data from the Monthly Retail Trade Survey,the Annual Retail Trade Survey, and administrative records. The industry group comprises establishments primarily engaged in retailing new and used vehicles.

based approach (MBA) – such as is implemented in TRAMO/SEATS[3] – by DFA methodology, as presented in Section 4. (Demetra+ is a program for seasonal adjustments that was developed and published by Eurostat European Commission; see `https://joinup.ec.europa.eu/software/demetraplus/description`. The package provides a user-friendly interface to TRAMO-SEATS and X-12-ARIMA.) We then successively refine the target signal, the spectral estimate and the MSE criterion by relying on methodology proposed in Section 4. Filter performances will be quantified in terms of MSE.

5.1.1 Replication of the Model-Based Approach by DFA

We propose results for data in *levels* (denoted X_t^{level}) as well as in *first differences* (denoted X_t^{diff}), but our emphasis on economic growth (or contraction) makes the use of differenced data natural. For sake of comparison all filter designs rely on *linearized* time series[4]; see Figure 1. The shaded regions in the figure highlight recessions as declared by the NBER.

Models identified by TRAMO for data in levels and in first differences are

$$\begin{aligned}(1-B)(1-B^{12})X_t^{\text{level}} &= (1-0.20136B)(1-0.61129B^{12})\epsilon_t^{\text{level}}\\ (1-B^{12})X_t^{\text{diff}} &= (1-0.63903B^{12})\epsilon_t^{\text{diff}},\end{aligned}$$

with $\{\epsilon_t\}$ denoting white noise. The first model (an airline model) selected by TRAMO reflects trend as well as seasonal features of auto-sales, in levels. As expected, the double unit root at frequency zero of the airline model reduces to a single root after differencing. Both models are deemed adequate according to the relevant diagnostics. Because the series are noisy, we decide to smooth them in our analysis, and therefore consider *trend* signals. Figure 2 plots (log-transformed) pseudo spectral densities and amplitude functions of canonical trends – symmetric and concurrent filters – for data in levels (top panel) and in first differences (bottom panel). That is, we set $\Psi(B)$ equal to the WK trend filter arising from the canonical decomposition (cf. Hillmer and Tiao (1982)) of each fitted model; this is the so-called canonical trend, and forms our target signal.

All spectral functions are generated by SEATS (based on the above time series model) and imported into the graphs. Note that SEATS generates *squared* gain functions (thus amplitude functions are obtained by the square root transformation). Log-transformed spectra can be negative-valued; singularities at the unit-root frequencies are truncated in the graphs.

Having specified (bi-infinite) signals, we now analyze real-time filters. Specifically, we demonstrate that DFA is able to replicate concurrent SEATS filters. For this purpose we insert the canonical trend $\Psi(z)$ (the target) and the differenced series' estimated spectral density $\overline{f}_{\widehat{\xi}}$ for g in (8), and obtain the transfer function $\widehat{\Psi}(z)$ by DFA. That is, we determine $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}$, the model-based

[3] All results in this article are based on version 1.0.2.2228 of Demetra+, downloaded on Feb.15, 2012.

[4] We check for logarithmic transformation, trading day effects, and outliers.

LPF described in Section 4. Here the class $\mathcal{G}$ of concurrent filters is taken to be appropriate restrictions of the large class $\mathcal{M}_{n-1}$, where n is the sample size. The necessary restrictions arise from the unit roots in the data process, and are given by equation (4); we write $\overline{\mathcal{M}}_{n-1}$ to denote this restricted class of filters.

This LPF can be compared to the semi-infinite concurrent filter based on the canonical decomposition – formulas for such can be found in Bell and Martin (2004). A comparison of SEATS and DFA in Figure 2 confirms that both curves are virtually indistinguishable, up to negligible finite-sample deviations due to the Gibbs phenomenon (Findley and Martin, 2006). This discrepancy is essentially due to our use of $\mathcal{M}_{n-1}$ in lieu of $\mathcal{M}_\infty$ (the maximal possible set of concurrent filters) in our computation of the model-based LPF.

5.1.2 MSE Performances for the Ideal Trend: MBA versus DFA

As we also want to emphasize economic growth, we now focus on the differenced series. Figure 3 compares the canonical trend of SEATS to the ideal low-pass trend (cf. Example 8) with $\mu = \pi/20$ (this particular specification in business-cycle analysis is further described at `http://www.idp.zhaw.ch/usri`.). The SEATS trend is very smooth, as expected from its transfer function (Figure 2, bottom panel). Note that all series are standardized for ease of visual inspection (otherwise, the output of the canonical trend would appear compressed). The finite sample symmetric ideal trend is a truncated version of the bi-infinite sample target; it cannot be computed towards the sample boundaries. Visual inspection suggests that the ideal trend is able to extract pertinent signals from the data. In particular, recessions are anticipated by steep downturns of the smoothed log-returns. In contrast, the canonical trend of SEATS seems to smooth out economic downturns. Because we are interested in tracking recession signals, our results suggest replacement of the model-based signal by the ideal trend in our LPP. This particular choice better reflects the purpose of our research, namely to identify economic expansion and contraction.

In this subsection we continue to focus upon the *differenced* auto-sales series. Having specified our new target signal $\Psi(z) = 1_{[-\pi/20,\pi/20]}(\lambda)$, we now compare the performances of two real-time LPFs: the model-based LPF (the MBA filter) relies upon the pseudo-spectral density derived from TRAMO, whereas the empirical LPF (called the DFA filter) uses the periodogram to address the estimation problem. In this context the MBA filter reflects a model-based estimate (using $\overline{f}_\xi$ for g) of a non-model-based target (the ideal low-pass). Note that our class of filters is still $\overline{\mathcal{M}}_{n-1}$, as in the previous subsection. The key difference is in the target, which itself is not a moving average filter (actually, $\Psi(B) \in \overline{\mathcal{M}}_\infty$), and can never arise as a Wiener-Kolmogorov signal extraction filter when using component ARIMA models. Both the empirical LPF and the model-based LPF utilize the DFA, but we will here refer to the former as the DFA filter and the latter as the MBA filter, to emphasize that models are being utilized to obtain the model-based LPF.

So estimates are obtained by inserting the common target, the ideal trend $\Psi(z)$, and the design-specific spectral estimates (either the SEATS spectrum $\overline{f}_{\widehat{\xi}}$ or the periodogram I) in equation (8) to obtain the LPFs $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(z)$ and $\widehat{\Psi}_{\theta(I)}(z)$, respectively the MBA and DFA filters. Because the TRAMO-SEATS model has (single) unit roots at frequency zero and at all seasonal frequencies, the MBA filter will be subject to first-order constraints at these frequencies[5] – see equation (4). In contrast, the DFA filter does not assume any unit-roots, which illustrates that we do not expect log-returns of sales of a physical good (autos) to be subject to asymptotically unbounded trends or unbounded seasonal fluctuations. These design decisions reflect different plausible interpretations of the same real-time estimation problem, and we expect the performance measures to reflect the marked idiosyncrasies of both designs. Figure 4 compares both real-time estimates (all filter outputs are similar in scale and in level, and therefore we omit unnecessary standardization, i.e., scales reflect original log-returns). Figure 5 provides evidence of the aforementioned design idiosyncrasies: the amplitude function of the MBA filter is equal to one at frequency zero and vanishes at the seasonal frequencies, whereas the DFA filter is unconstrained. We next analyze the performances of the proposed designs in order to get a better understanding of their potential.

Given the model-based LPF $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(B)$ and the empirical LPF $\widehat{\Psi}_{\theta(I)}(B)$, it is of interest to gauge their performance in terms of both the model-based estimate of the spectrum, and the empirical estimate. In the case of the model-based LPF, we assume the presence of unit roots, as discussed in the preceding paragraphs, and hence we utilize $\overline{f}_{\widehat{\xi}}$ for g in equation (8), where $\delta(z) = 1 - z^{12}$. But for the empirical LPF we substitute I for g and have no unit roots, so that $\delta(z) = 1$. The integrand of (8) in the case of the minimal MSE model-based LPF is

$$\frac{\left|\Psi(z) - \widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(z)\right|^2}{|\delta(z)|^2}\, \overline{f}_{\widehat{\xi}}(\lambda). \tag{9}$$

That is, any other choice of $\widehat{\Psi}(B) \in \overline{\mathcal{M}}_{n-1}$ must yield a higher MSE than the integral of (9). However, if we are to assess the performance of this model-based LPF on the actual auto-sales data (assuming the presence of unit roots $\delta(B) = 1 - B^{12}$), the LPP MSE is approximately given by the integral of

$$\left|\Psi(z) - \widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(z)\right|^2 I(\lambda). \tag{10}$$

Here the periodogram I is for the log-return (not seasonally adjusted) data; recall from Section 2 that the LPP MSE for nonstationary processes will have a unit root factor $|\delta(z)|^2$ in the denominator.

On the other hand, the empirical LPF (which again, assumes no unit roots are present) yields the minimal MSE of (8), which will have integrand

$$\left|\Psi(z) - \widehat{\Psi}_{\theta(I)}(z)\right|^2 I(\lambda). \tag{11}$$

[5]The amplitude equals one in frequency zero and it vanishes in seasonal frequencies.

Any other choice of $\widehat{\Psi}(B) \in \mathcal{M}_{n-1}$ must yield a higher MSE than the integral of (11). Note that $\widehat{\Psi}(B)$ need not be in $\overline{\mathcal{M}}_{n-1}$, and $\mathcal{M}_{n-1} \supset \overline{\mathcal{M}}_{n-1}$. The periodogram in (11) used to obtain the DFA filter is the same as in (10). It may also be of interest to examine the empirical LPF when a model-based estimate of the spectrum is substituted for the periodogram. Because $\overline{f}_{\widehat{\xi}}$ is determined from seasonally differenced log-returns, the pseudo-spectral density $\overline{f}_{\widehat{\xi}}(\lambda)/|1-z^{12}|^2$ should be used as our model-based estimate; this plays the role of g in (8) with $\delta(z) = 1$. This yields

$$\left|\Psi(z) - \widehat{\Psi}_{\theta(I)}(z)\right|^2 \frac{\overline{f}_{\widehat{\xi}}(\lambda)}{|1-z^{12}|^2} \tag{12}$$

for the integrand. Note that because (4) need not be satisfied for the unit roots of $1-z^{12}$, it is possible that (12) is unbounded and non-integrable. In our particular implementation this was the case, with the result that the integral of (12) is infinite.

The four quantities described above are reported in the first column of table 1, with equations (9), (10), (11), and (12) describing each of the four rows. In contrast, the Time-Domain MSEs reported in the second column of the table are obtained by computing the time-domain empirical MSE between the target (utilizing a finite sample approximation to the ideal trend) and the real-time estimates: note that these numbers depend on the filter design (MBA vs. DFA) only. Because of its symmetry, the target filter cannot be computed towards the ends of the sample. Moreover, the finite sample reference signal is a truncated version of the true bi-infinite target signal. In contrast, the frequency domain measures in the first column do not rely on filter outputs; they are *full-sample* estimates with respect to the *true* (bi-infinite) ideal trend. One has to keep these distinctions in mind when interpreting reported numbers, because the Great Recession affects performance measures differently in accordance with the sample period under scrutiny.

In order to gauge the importance of the Great Recession on performance, we also report *adjusted* Time-Domain MSEs (in parentheses) in the second column of table 1 for a period prior to the start of the recession; as can be seen, adjusted numbers from this truncated sample are one third to one quarter of the unadjusted MSEs. Given the striking impact of the Great Recession, we tend to interpret absolute numbers with caution[6]. Nonetheless much useful information regarding the virtues of MBA or DFA can be extracted from the table. We first fix attention on the Frequency-Domain MSEs in the first column of table 1: estimates for MBA ($2.02e-05$) and DFA ($1.63e-05$) based on the periodogram (rows 2 and 4) suggest that DFA outperforms MBA by a reduction of MSE by approximately 24.5%. As noted above, inserting the SEATS spectrum in the case of DFA (row 3) is not meaningful because DFA ignores unit-roots of the model. Interestingly, the model-based MSE in row 1 ($4.72e-06$) is markedly smaller than the periodogram based MSE ($2.02e-05$).

[6]If the Great Recession is a rare event, then we would be better advised to report adjusted numbers only (those in parentheses); otherwise, combining pre-recession figures with recession numbers would be legitimate. Because this decision is to some extent a matter of taste, we publish both numbers.

A comparison with the Time-Domain MSEs in the second column may help to alleviate this conflict: periodogram-based MSEs emphasize unadjusted MSEs, whereas the model-based estimate seems to comply with recession adjusted numbers (in parentheses). It is as if the model ignored the singular event of the Great Recession (possibly because the innovation variance has been adjusted for recession outliers). Adjusted Time-Domain MSEs ($7.04e-06$ for MBA and $5.17e-06$ for DFA) confirm and exceed the previous efficiency gain by DFA (36.1%). Finally, unadjusted Time-Domain MSEs ($2.21e-05$ for MBA and $2.06e-05$ for DFA) confirm the dominance of DFA, once again, but to a lesser extent[7]. In the latter case, the mean square aggregate is unable to distinguish the marked design peculiarities over the available (truncated) history due to a singular event, which balances out pros and cons in a more or less fortuitous way.

	Frequency-Domain MSE's	Time-Domain MSE's
MBA/SEATS	4.72e-06	2.21e-05 (7.04e-06)
MBA/Periodogram	2.02e-05	2.21e-05 (7.04e-06)
DFA/SEATS	Inf	2.06e-05 (5.17e-06)
DFA/Periodogram	1.63e-05	2.06e-05 (5.17e-06)

Table 1: Frequency-Domain and Time-Domain MSEs for MBA and DFA concurrent filters targeting the ideal low-pass trend, when applied to log-returns of auto-sales.

To conclude, we gauge both filters with respect to amplitude and time-shift functions as plotted in Figure 5. The time-shift is defined by $\widehat{\Phi}_{DFA/MBA}(\omega)/\omega$, where $\widehat{\Phi}_{DFA/MBA}(\omega)$ is the phase function of either design, and $\omega \in [0, \pi]$. Positive numbers indicate that trigonometric signals would be shifted by a corresponding number of months to the future by the filter (i.e., this corresponds to a delay). A smaller time-shift of the DFA filter in the pass-band (around frequency zero) is a desirable property, because turning-points of the series could be detected earlier. As an example, the expansion following the trough of the Great Recession is affirmed earlier by the DFA, and similarly for the onset of the dotcom recession, whose peak is timed earlier by DFA – see Figure 4. The stronger noise rejection of DFA (the amplitude function of the DFA filter is closer to zero in the stop-band) is also an advantage, as compared to MBA, because log-returns are noisy. In contrast, MBA outperforms DFA with respect to amplitude characteristics in the pass-band, since the unit-root constraint imposes a perfect match at frequency zero. Based on this contrast, we may infer that MBA is likely to outperform DFA if the transformed series is subject to a permanent and slowly varying level-shift; preferences for either filter design may be adopted, depending on whether log-

[7]As previously noted, the Great Recession affects performances differently according to the sample period: Frequency-Domain MSEs emphasize the full sample, while adjusted Time-Domain MSEs (in parentheses) emphasize a period prior to the Great Recession and unadjusted MSEs are in-between (because of truncation of the target signal – see Figure 4).

returns of the data are deemed to follow such a non-stationary pattern[8]. A comprehensive treatment of frequency-domain characteristics is provided in Wildi and McElroy (2014), where the authors derive a generic optimization criterion addressing noise-suppression, timeliness, and accuracy of real-time designs.

5.2 Seasonal Adjustment: MBA vs. DFA

To illustrate the scope of the proposed DFA we here propose to study real-time seasonal adjustment of MidWest housing starts. We study "New Residential Construction, 1964-2012, Housing Units Started, Single Family Units" from the Survey of Construction of the U.S. Census Bureau, available at `http://www.census.gov/construction/nrc/how_the_data_are_collected/soc.html`. This series, along with the three other major regional starts series for the U.S., have great importance for understanding the U.S. economy – both retail and housing are key facets of consumption and production activity in advanced economies. The MidWest series is impacted by winter weather more heavily than the South and West regional series, but is similar to the NorthEast regional series in this aspect; this data feature makes modeling the seasonal pattern more challenging, and some authors have even advocated seasonally heteroscedastic models (Trimbur and Bell, 2012). Original and log-transformed data are shown in Figure 6. In either case, that is with or without log-transformation, the seasonality appears to have changing amplitude (other Box-Cox transforms could have been selected). This type of behavior is frequently encountered in the practice of seasonal adjustment – more interesting, and potentially challenging, is the variable strength of seasonal frequencies exhibited in the periodogram, as described below.

5.2.1 Model-Based Filter

TRAMO selects a logarithmic transformation for the MidWest starts (MW henceforth). The following airline model was identified for the transformed series $\{X_t\}$:

$$(1-B)(1-B^{12})X_t = (1-0.440B)(1-0.845B^{12})\epsilon_t.$$

Diagnostic statistics detect seasonal instability, as was to be expected, but otherwise model residuals pass the usual checks. The model-based (pseudo-) spectral density as well as the periodogram of $\{X_t\}$ are compared in Figure 7: we use logarithmic transforms in order to highlight the salient features in the data. The original and log transformed periodograms suggests that the seasonal pattern is more complex than the model can possibly capture: the first two or three seasonal peaks (at frequencies $\pi/6$, $2\pi/6$, and $3\pi/6$) clearly dominate. In particular, these first peaks appear to be wider than those generated by the model. In contrast, the last three peaks (at frequencies $4\pi/6$, $5\pi/6$, and $6\pi/6$) are either non-existent or negligible. In this particular situation, characterized

[8]We find this assumption rather implausible since, mean-reversion seems to apply to both recession episodes.

by an inhomogeneous seasonal pattern, the model seems to adopt a compromise, whereby the importance of the first three peaks is understated and the presence of the last three peaks is exaggerated. In fact, the airline model relies on a single parameter, the seasonal MA coefficient (of value .845), to fit the nuanced seasonal pattern; the entanglement of the various spectral peaks impedes the model's flexibility. More nuanced models are possible, such as the stochastic cycle representation of seasonality (this provides a parameter for each of the six seasonal frequencies) described in Harvey (1989) and Proietti and Grassi (2012), or the generalized Airline model (Aston et al., 2007). Our objective in this illustration is not to defeat a "straw man" MBA competitor of the DFA, but rather to highlight the distinctive features of both approaches.

The model-based gain functions of the SA filters are shown in Figure 8: the symmetric bi-infinite target (solid line), the concurrent semi-infinite SEATS filter (small dots), and the concurrent finite-length DFA-replication (dotted) are compared. Both infinite-length filters are generated by SEATS[9] whereas the finite-length filter (length 120, or 10 years) is generated by the DFA. The latter filter replicates the one-sided infinite filter of SEATS, up to well-known finite-sample approximation errors. Increasing the length of the finite DFA-filter further would improve the approximation up to arbitrarily small deviations, but signal extraction performances would not improve; therefore we may restrict attention to finite filters of length 120 (10 years). As expected, the seasonal dips of the filters follow a uniform pattern with nearly constant width (cf. the upper panels of Figure 7).

5.2.2 DFA: new Target and Periodogram

We next modify the target by substituting an ideal (bi-infinite) SA target to the model-based bi-infinite SA filter. We first keep the SEATS-spectrum fixed and then we also modify the spectrum, replacing the model-based estimate by the periodogram in the DFA criterion (8). Figure 9 plots the new target specification and the resulting real-time DFA filters based on the SEATS-spectrum (dotted) and on the periodogram (dashed) together with the periodogram of $\{X_t\}$. The target SA filter is deliberately simple: it is almost an identity, except for three seasonal dips at the dominant peaks $\pi/6$, $2\pi/6$, and $3\pi/6$, where the function vanishes exactly. Note that the dips are slightly wider than is true of the model-based target of the previous section, because the actual dominant peaks are wider than allowed for by the model[10]. Obviously, this generic target specification could account for undesirable spectral peaks of arbitrary width and of arbitrary location in a time series; for example, non-seasonal calendar effects could be accounted for in the same vein. The amplitude functions of the real-time filters show evidence of finite-sample ripples in the vicinity of the dips,

[9] We applied the square-root transformation to the original output, since SEATS generates squared gain functions by default.

[10] The signal-specification can be automated, depending only on the periodogram to specify the width of potential seasonal peaks.

as is to be expected (due to the Gibbs phenomenon[11]), but potentially undesirable effects are negligible; on the contrary, (real-time) seasonal adjustment performances tend to improve on the original model-based approach, as shown in the next section.

5.2.3 Comparison of Real-Time SA-Filters

We apply the finite MBA-filter of section 5.2.1 and the finite DFA-filters of the previous subsection (both of length 10 years) to $\{X_t\}$, and analyze the resulting outputs. Specifically, Figure 10 compares the periodograms of the filtered series. In order to highlight the relevant characteristics of the filters, we have split the frequency band, omitting the dominant trend frequencies, emphasizing either the first dominant seasonal peaks (upper panel) or the remaining negligible seasonal peaks (bottom panel). The DFA filter based on the periodogram (shaded) damps the dominant peaks in $\pi/6$ and $2\pi/6$ the most effectively, followed by the DFA filter based on SEATS' spectrum (dotted), and lastly the original model-based approach (solid). This ranking was to be expected, since the spectral peaks in SEATS' spectrum are narrower at the dominant frequencies. By design the last three peaks $4\pi/6$, $5\pi/6$ and $6\pi/6$ remain unaffected, since our target function does not dip in these frequencies. Note that we could have specified spectral dips of arbitrary width at any seasonal or non-seasonal frequency in the DFA target of the previous subsection. We refrained from doing so, partly because the magnitude of the last three peaks is negligible, and partly because we wanted to illustrate the scope and the flexibility of our procedure. Indeed, a particular actualization of the real-time filter can be obtained very easily in the DFA by specifying a corresponding target in equation (8). The resulting facility and flexibility are obtained by addressing filter-coefficients directly in the generalized DFA criterion.

6 Conclusion

Real-time signal extraction is a topic of considerable applied interest in macroeconomics and finance. Whether the application is forecasting, seasonal adjustment, business cycle analysis, trend estimation, or turning point identification, there is a market-driven need to obtain real-time extractions that are both timely and accurate. This paper focuses on matching the particular real-time filter to the objectives of the practitioner through the formalism of a Linear Prediction Problem (LPP). Real-time filters can then be designed as analytic solutions that solve a given LPP, and further can be approximated either through a class of time series models or through a suitable class of concurrent filters. The latter approach, which uses the periodogram to "fit" a parametrized filter to the data, is called the Direct Filter Approach (DFA), and can be contrasted with the former

[11] Use of the periodogram alleviates the Gibbs phenomenon to some extent, because the spectral peaks are "less sharp" than with SEATS' spectrum.

Model Based Approach (MBA), which relies upon a specified model being an accurate portrait of the process' dynamics.

The three main contributions of the paper, which we believe to be novel and useful, are: (1) we define and solve LPPs, providing several key examples; (2) we treat model-fitting via LPP minimization, and describe the resulting model-based real-time filters; (3) we connect LPPs to the DFA, and describe the resulting real-time filters. We show that DFA is broader than model-based approaches, and can yield improved performance in cases where a good model is hard to identify. Our treatment is illustrated through two main examples: trend estimation from a retail series, and seasonal adjustment of a construction series. Other work – the subject of the companion paper Wildi and McElroy (2014) – further explores the design of filters, taking into account their frequency domain properties (described via the gain and phase delay functions) directly in the DFA criterion. Other extensions, such as multivariate filtering, are also under investigation.

In order to encourage other scientists to understand and utilize our work, this paper has been generated via SWEAVE, and can be recomputed by following instructions on the Internet – see the Appendix. The first author's blog contains links to code and frequent updates to the ongoing process of discovery. The DFA paradigm is currently being utilized, in various incarnations, to address real-time signal extraction problems in economics and finance, in Switzerland and other countries. Whereas some other methodologies also offer tuning parameters to adjust real-time filters, we believe that our formulation is the most direct and intuitive, and moreover can be made model-free. This feature can be an advantage, when a scientist is concerned that forecasts or extractions may be unduly restricted via their generation through the modeling "prism"; yet for data that truly warrants a particular model, the LPP can be fitted so as to yield the most appropriate parameter choices for the given time series. We believe this flexibility and power to be compelling facets of this paper's methodology, and it is our hope that the readers will utilize the code to analyze new and diverse applications.

Acknowledgements. The second author thanks the Institute of Data Analysis and Process Design (IDP-ZHAW) for hosting a visit that facilitated the research. The first author benefitted from a Summer at Census grant. We thank Christopher Blakely for stimulating comments and discussion on this work.

Appendix

A.1 Proofs of Results

Proof of Proposition 1. In order for a solution to be optimal, it is sufficient that the resulting error process be uncorrelated with the data $X_{t:}$, because this guarantees that the solution is the Gaussian conditional expectation. (For non-Gaussian processes, optimality refers to minimum MSE

among linear estimates, and the same criteria are in force – see Bell (1984) for background.) If we can show that the real-time signal extraction error process depends only on future innovations, then using (1) it must be uncorrelated with $X_{t:}$, establishing optimality. This logic utilizes the assumption that the initial values are uncorrelated with the innovations. The filter error of the putative solution is $\Delta(z) = \Psi(z) - \widehat{\Psi}(z)$, which is given by

$$\begin{aligned}
\Delta(z) &= \sum_{h<0} \psi_h \left(z^h - \sum_{j=1}^{d} A_{j,d-h} z^{d-j} - \sum_{k=1}^{-h} \xi_{-h-k} [\Pi]_k^{\infty}(z) z^{-k} \delta(z) \Pi^{-1}(z) \right) \\
&= \sum_{h<0} \psi_h \left(\sum_{k=0}^{-h-1} \xi_k z^{k+h} - \sum_{k=0}^{-h-1} \xi_k [\Pi]_{-h-k}^{\infty}(z) z^{k+h} \Pi^{-1}(z) \right) \delta(z) \\
&= \sum_{h<0} \psi_h \sum_{k=0}^{-h-1} \xi_k z^k [\Pi]_0^{-h-k-1}(z) z^h \Pi^{-1}(z) \delta(z) \\
&= \sum_{h<0} \psi_h [\Pi/\delta]_0^{-h-1}(z) z^h \Pi^{-1}(z) \delta(z).
\end{aligned}$$

The second equality uses (2) and a change of index variable. The fourth equality uses another algebraic relation, first established in McElroy and Findley (2010), that $\sum_{k=0}^{-h-1} \xi_k B^k [\Psi]_0^{-h-k-1}(B) = [\Psi/\delta]_0^{-h-1}(B)$. Now this algebra yields $\tau(B)$ explicitly (i.e., $\delta(B)$ divides $\Delta(B)$), namely

$$\tau(B) = \sum_{h>0} \psi_{-h} [\Pi/\delta]_0^{h-1}(B) F^h \Pi^{-1}(B).$$

Hence the real-time error process is $\tau(B) W_t = \sum_{h>0} \psi_{-h} [\Pi/\delta]_0^{h-1}(B) \epsilon_{t+h}$, which is clearly a linear function of future innovations $\epsilon_{t+1}, \epsilon_{t+2}, \cdots$. This completes the proof. □

Proof of Theorem 1. Note that $\omega(g)$ is a minimizer of $J_\Psi(\omega, g)$, so we can do a Taylor series expansion of the gradient at $\omega(I)$ and $\omega(\widetilde{f})$. This yields the asymptotic expression (cf. Taniguchi and Kakizawa (2000))

$$\sqrt{n} \left(\omega(I) - \omega(\widetilde{f}) \right) = o_P(1) - H^{-1}(\omega_{\widetilde{f}}) \, \frac{\sqrt{n}}{2\pi} \int_{-\pi}^{\pi} \nabla r_{\omega(\widetilde{f})}(\lambda) \left(I(\lambda) - \widetilde{f}(\lambda) \right) d\lambda,$$

where r_ω is defined in the theorem. Our assumptions allow us to apply Lemma 3.1.1 of Taniguchi and Kakizawa (2000) to the right hand expression above, and the stated central limit theorem is obtained. □

Proof of Theorem 2. This is proved in the same exact manner as Theorem 1.

Proof of Theorem 3. First, convergence in probability follows from the continuity of L. The Central Limit Theorem follows from the delta method:

$$\sqrt{n} \left(\theta(f_{\widehat{\xi}}) - \theta(f_{\widetilde{\xi}}) \right) = \nabla L(\widetilde{\xi}) \, \sqrt{n} \left(\widehat{\xi} - \widetilde{\xi} \right) + o_P(1)$$

by Taylor series expansion of L; then use the known CLT for $\widehat{\xi}$. □

A.2 R-Code

DFA and its multivariate version MDFA are extensively discussed on the Signal-Extraction and Forecasting (SEF) blog `http://blog.zhaw.ch/idp/sefblog`. The following entries are potentially relevant in the context of this article:

- The Sweaved version of this paper is available under `http://blog.zhaw.ch/idp/sefblog/index.php?/archives/380-Optimal-Real-Time-Filters-for-Linear-Prediction-Problems.html`: Interested users can download the R-code and replicate results in this paper.
- Trilemma paper: `http://blog.zhaw.ch/idp/sefblog/index.php?/archives/381-The-Trilemma-Betwee-Timeliness-and-Smoothness-in-Real-Time-Signal-Extraction.html`
- Replication and customization of the Hodrick-Prescott filter: `http://blog.zhaw.ch/idp/sefblog/index.php?/archives/261-Replication-of-the-Hodrick-Prescott-Filter-by-I-MDFA.html`
- Replication and customization of the Christiano-Fitzgerald filter: `http://blog.zhaw.ch/idp/sefblog/index.php?/archives/264-Replication-of-the-Christiano-Fitzgerald-CF-Filter-by` `html`

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] Aston, J., Findley, D., McElroy, T., Wills, K., and Martin, D. (2007) New ARIMA models for seasonal time series and their application to seasonal adjustment and forecasting. U.S. Census Bureau Research Report RRS2007/14.

[3] Banbura, M., Giannone,D., Reichlin,L. (2010) Nowcasting. *Working Paper Series No. 1275, European Central Bank.*

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

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

[6] Bell, W. and Hillmer, S. (1984) Issues involved with the Seasonal Adjustment of Economic Time Series. *Journal of Business and Economic Statistics* **2**, 291–320.

[7] 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.

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

[9] Cox, D. (1961) Test of separate families of hypotheses. In "Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability," Vol. 1. Berkeley: University of California Press, 105–123.

[10] Cox, D. (1962) Further results on tests of separate families of hypotheses. *Journal of the Royal Statistical Society, Series B* **24**, 406–424.

[11] 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.

[12] Dahlhaus, R., and Wefelmeyer, W. (1996) Asymptotically optimal estimation in misspecified time series models. *Ann. Statist.* **16**, 952–974.

[13] Findley, D. and Martin, D. (2006) Frequency domain analyses of SEATS and X-11/12-ARIMA seasonnal adjustment filters for short and moderate-length time series. *Journal of Official Statistics* **22**, 1–34.

[14] Findley, D., Monsell, B., Bell, W., Otto, M., and Chen, B. (1998) New capabilities and methods of the X-12-ARIMA seasonal adjustment program. *J. Bus. Econ. Stat.* **16**, 127–177.

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

[16] Haykin, S. (1996) *Adaptive Filter Theory.* Upper Saddle River, New Jersey: Prentice Hall.

[17] Hillmer, S. and Tiao, G. (1982) An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association* **77**, 63–70.

[18] Hodrick, R. and Prescott, E. (1997) Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit, and Banking* **29**, 1–16.

[19] Holan, S. and McElroy, T. (2012) "Bayesian seasonal adjustment of long memory time series." In *Economic Time Series: Modeling and Seasonality*, eds. Bell, W., Holan, S., McElroy, T. CRC Press. Boca Raton, FL.

[20] Hosoya, Y. and Taniguchi, M. (1982) A central limit theorem for stationary processes and the parameter estimation of linear processes. *Ann. Statist.* **10**, 132–153.

[21] Koopman, S., Harvey, A., Doornik, J., and Shepherd, N. (2000) *Stamp 6.0: Structural Time Series Analyser, Modeller, and Predictor.* London: Timberlake Consultants.

[22] Ladiray, D. and Quenneville, B. (2001) *Seasonal Adjustment with the X-11 Method.* New York: Springer.

[23] Maravall, A. and Caparello, G. (2004) Program TSW: Revised Reference Manual. Working paper 2004, Research Department, Bank of Spain. http://www.bde.es.

[24] 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.

[25] McElroy, T. (2008). Exact Formulas for the Hodrick-Prescott Filter. *Econometrics Journal* **11**, 1–9.

[26] McElroy, T. (2010). A Nonlinear Algorithm for Seasonal Adjustment in Multiplicative Component Decompositions. *Studies in Nonlinear Dynamics and Econometrics* **14**: No. 4, Article 6.

[27] McElroy, T. (2011). A Nonparametric Method for Asymmetrically Extending Signal Extraction Filters. *Journal of Forecasting* **30**, 597–621.

[28] McElroy, T. and Findley, D. (2010). Discerning Between Models Through Multi-Step Ahead Forecasting Errors. *Journal of Statistical Planning and Inference* **140**, 3655–3675.

[29] McElroy, T. and Holan, S. (2009). A Local Spectral Approach for Assessing Time Series Model Misspecification. *Journal of Multivariate Analysis* **100**, 604–621.

[30] McElroy, T. and Wildi, M. (2010). Signal Extraction Revision Variances as a Goodness-of-Fit Measure. *Journal of Time Series Econometrics* **2**, Iss. 1, Article 4.

[31] McElroy, T. and Wildi, M. (2013) Multi-Step Ahead Estimation of Time Series Models. *International Journal of Forecasting* **29**, 378–394.

[32] Proietti, T. and Grassi, S. (2012) Bayesian stochastic model specification search for seasonal and calendar effects. In W. Bell, S. Holan, and T. McElroy (Eds.), *Economic Time Series: Modeling and Seasonality.* New York: Chapman and Hall.

[33] Self, S. and Liang, K. (1987) Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association* **82**, 605–610.

[34] Taniguchi, M. and Kakizawa, Y. (2000) *Asymptotic Theory of Statistical Inference for Time Series.* Springer-Verlag, New York.

[35] 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.

[36] Trimbur, T. and Bell, W. (2012) Seasonal heteroscedasticity in time series data: modeling, estimation, and testing. In W. Bell, S. Holan, and T. McElroy (Eds.), *Economic Time Series: Modeling and Seasonality.* New York: Chapman and Hall.

[37] Wildi, M. (2005) Signal Extraction: Efficient Estimation, Unit-Root Tests and Early detection of Turning Points. *Lecture Notes in Economics and Mathematical Systems, 547: Springer.*

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

[39] Wildi, M. and McElroy, T. (2014) The Trilemma between Accuracy, Timeliness and Smoothness in Real-Time Signalextraction. `http://blog.zhaw.ch/idp/sefblog/index.php?/archives/381-The-Trilemma-Between-Accuracy,-Timeliness-and-Smoothness-in-Real-Time-Signal-Extraction.html` .

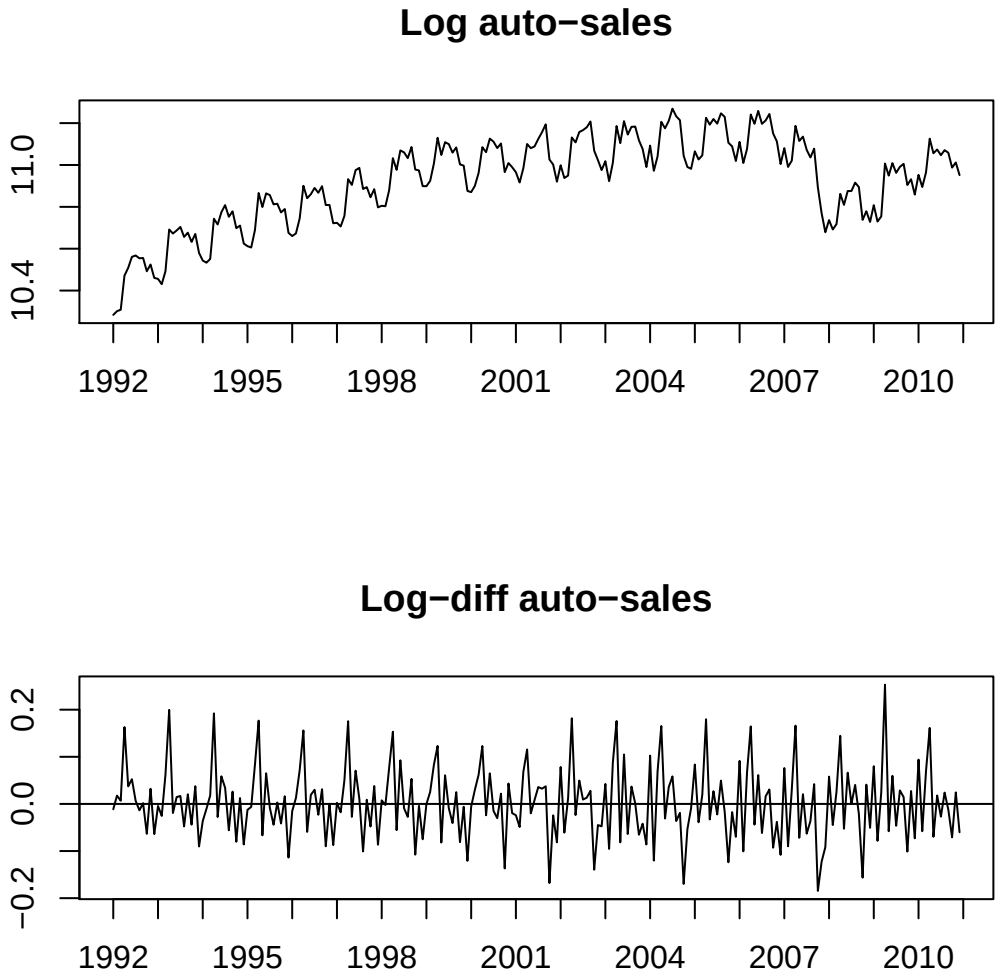

Figure 1: Linearized auto-sales series: levels (top) and first differences (bottom). Dotcom- and great-recession appear shaded.

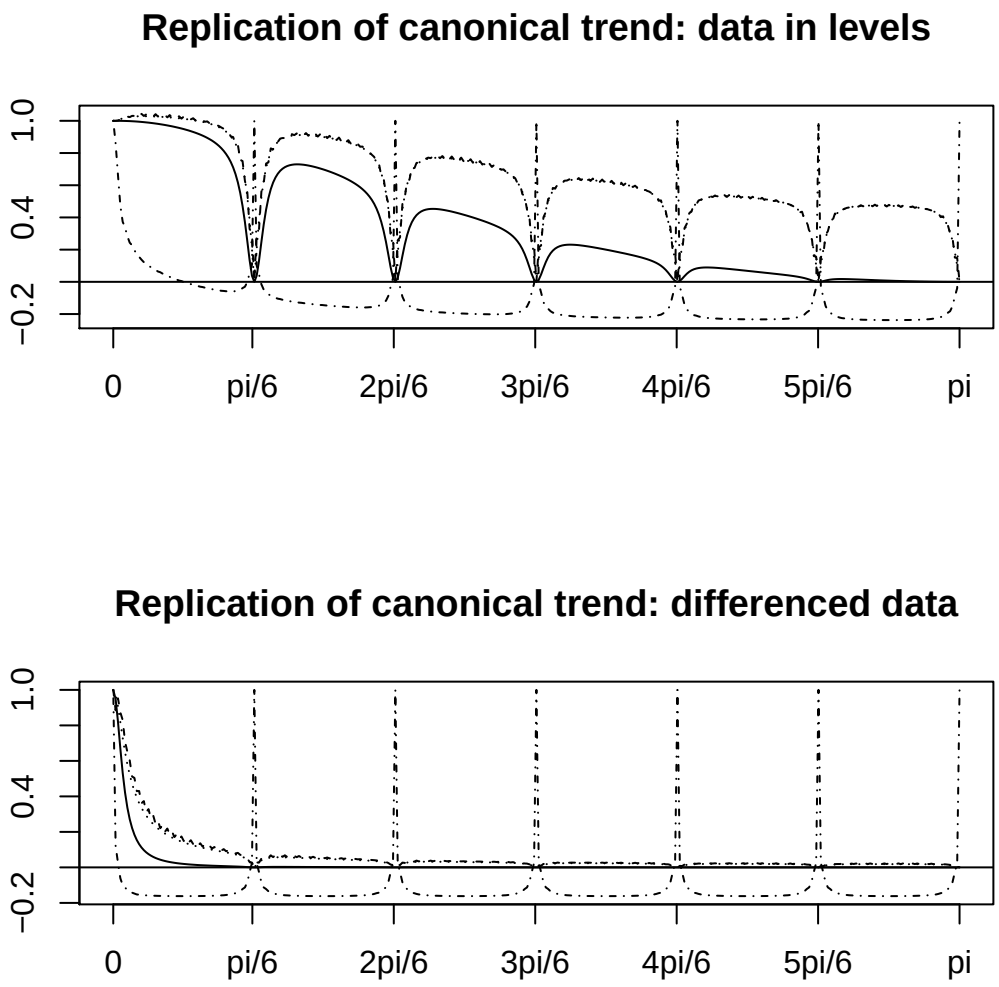

Figure 2: Model-based (log-transformed) pseudo-spectral densities (dot-shaded) and canonical trends: symmetric target filters (solid) and concurrent model-based filters (dotted), as well as DFA replications (shaded) are displayed, for the auto-sales in levels (top panel) and in first differences (bottom panel).

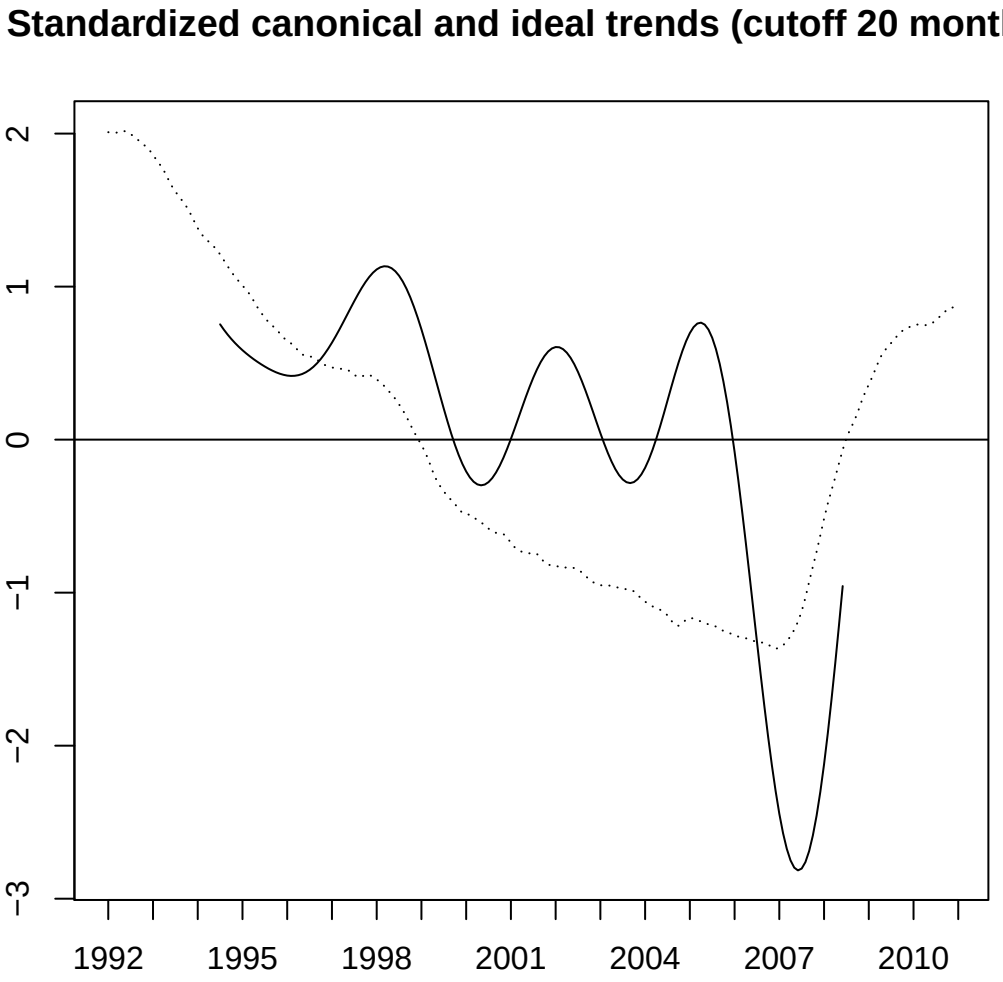

Figure 3: Ideal low-pass trend (solid) and canonical trend (dotted), both standardized, for auto-sales series in first differences.

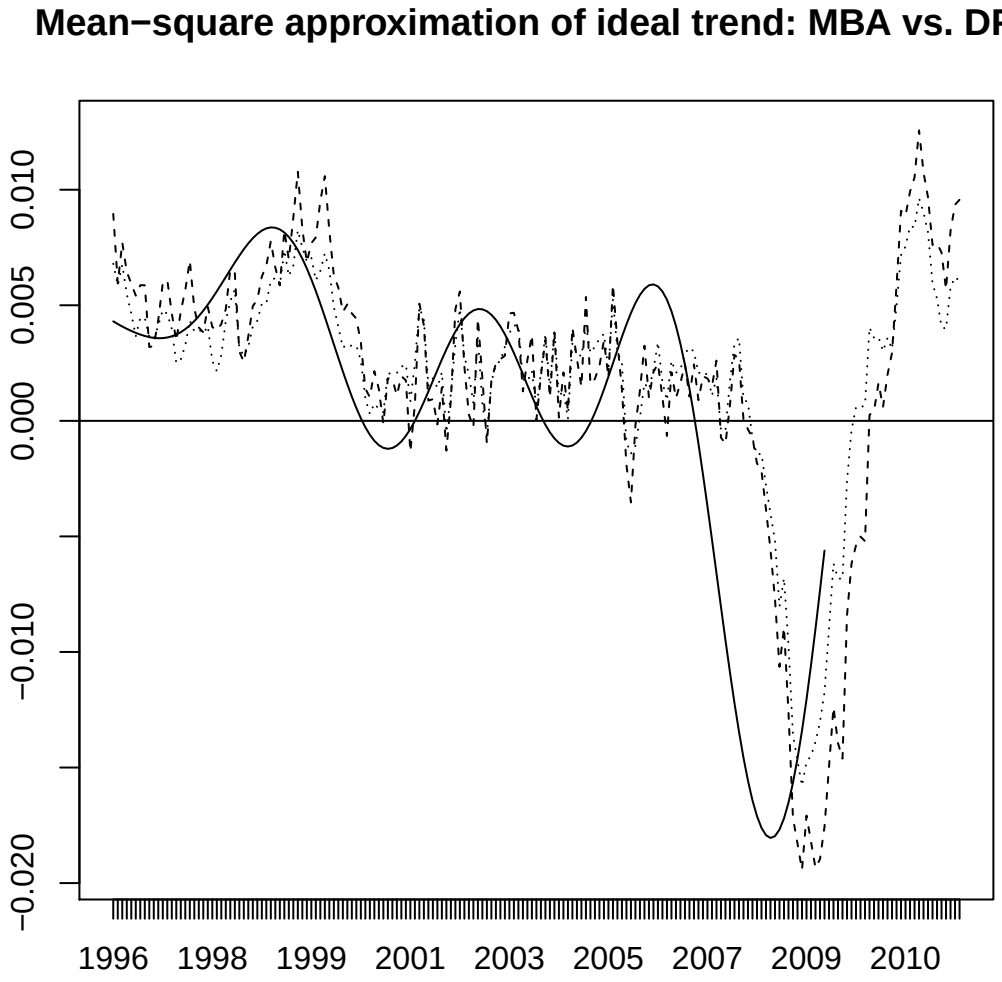

Figure 4: Real-time approximations of the ideal trend: MBA filter shaded (model-based GPF $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(B)$) versus DFA filter dotted (empirical GPF $\widehat{\Psi}_{\theta(I)}(B)$) applied to auto-sales; ideal trend solid.

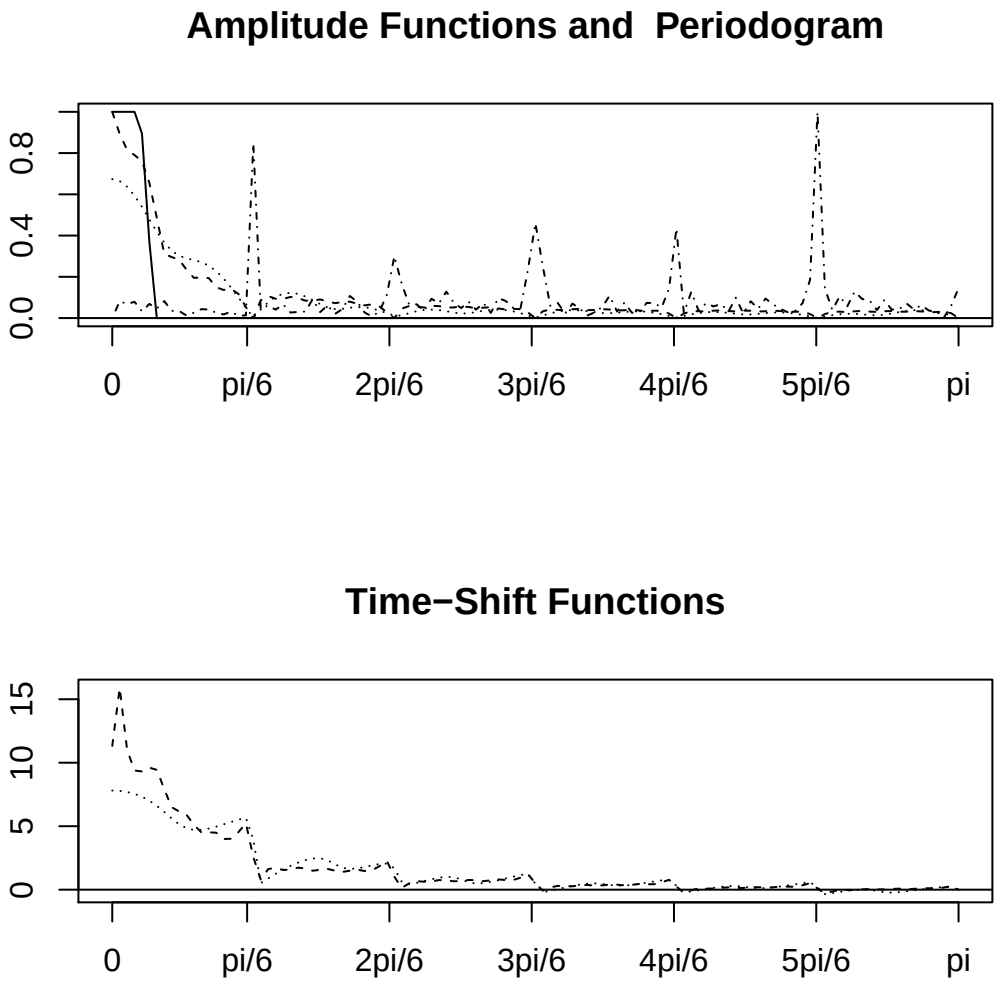

Figure 5: Periodogram of auto-sales log-returns (dot-shaded) and amplitude functions (top-graph) as well as time-shifts (bottom graph): MBA filter shaded (model-based GPF $\widehat{\Psi}_{\theta(\overline{f}_{\widehat{\xi}})}(z)$) versus DFA filter dotted (empirical GPF $\widehat{\Psi}_{\theta(I)}(z)$); ideal trend solid.

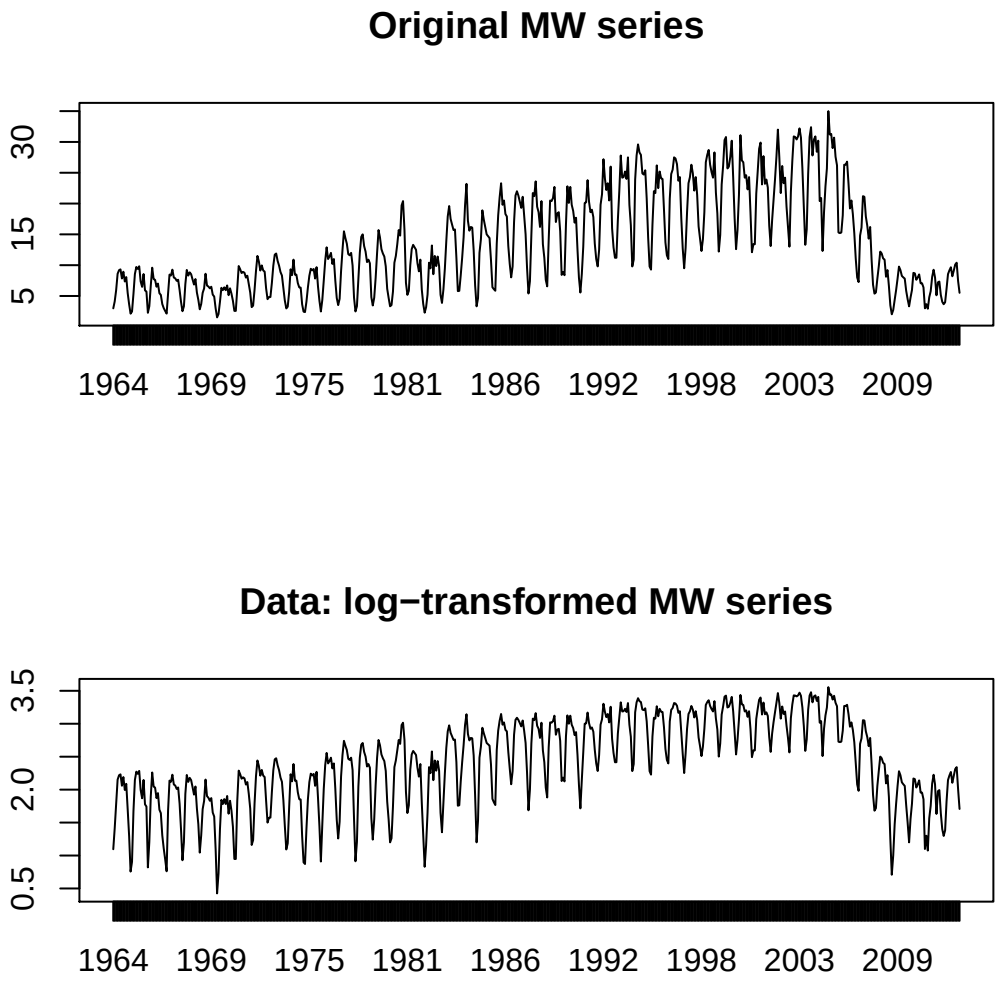

Figure 6: Original and log-transformed MW series.

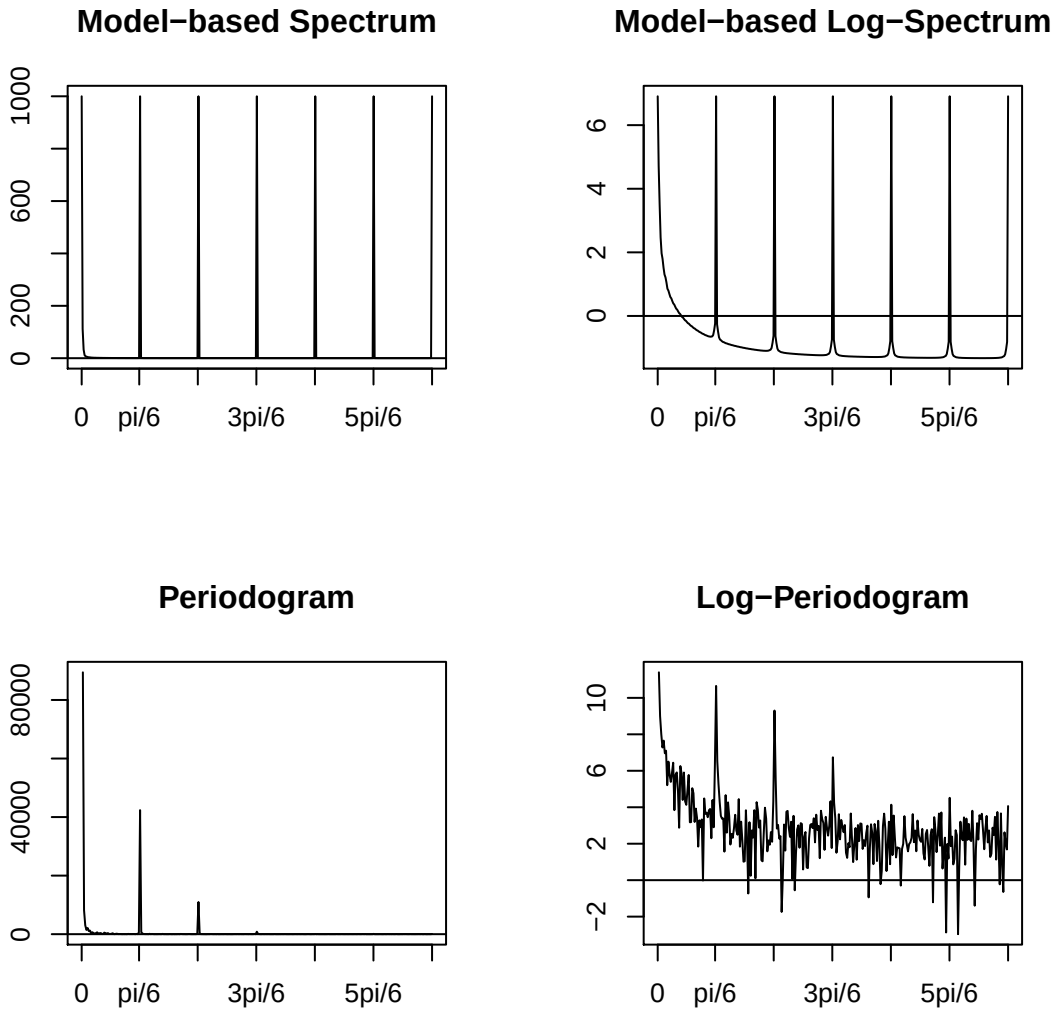

Figure 7: Spectral density estimates for the MW series. Original and log-transformed model-based spectrum (top) and periodogram (bottom).

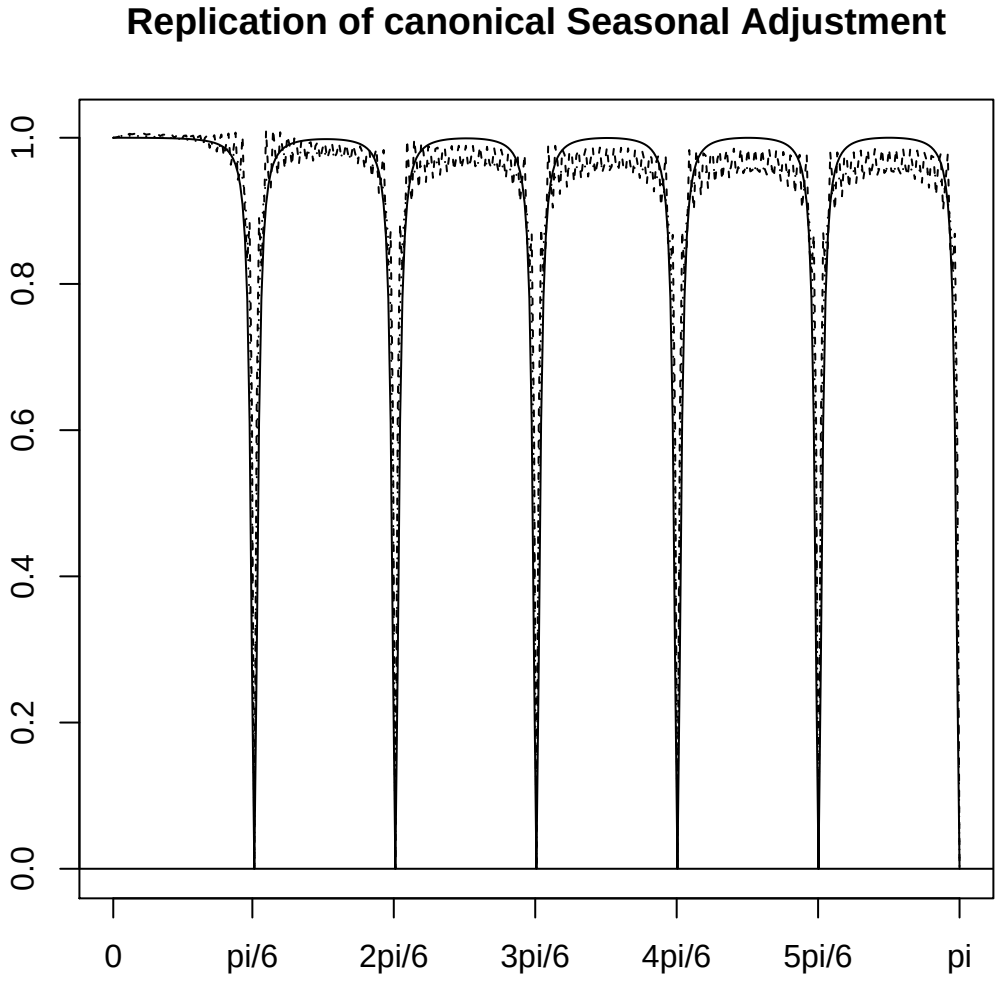

Figure 8: Gain functions of bi-infinite symmetric target (solid), semi-infinite concurrent (dotted), and finite concurrent (shaded) SA filters for the MW series.

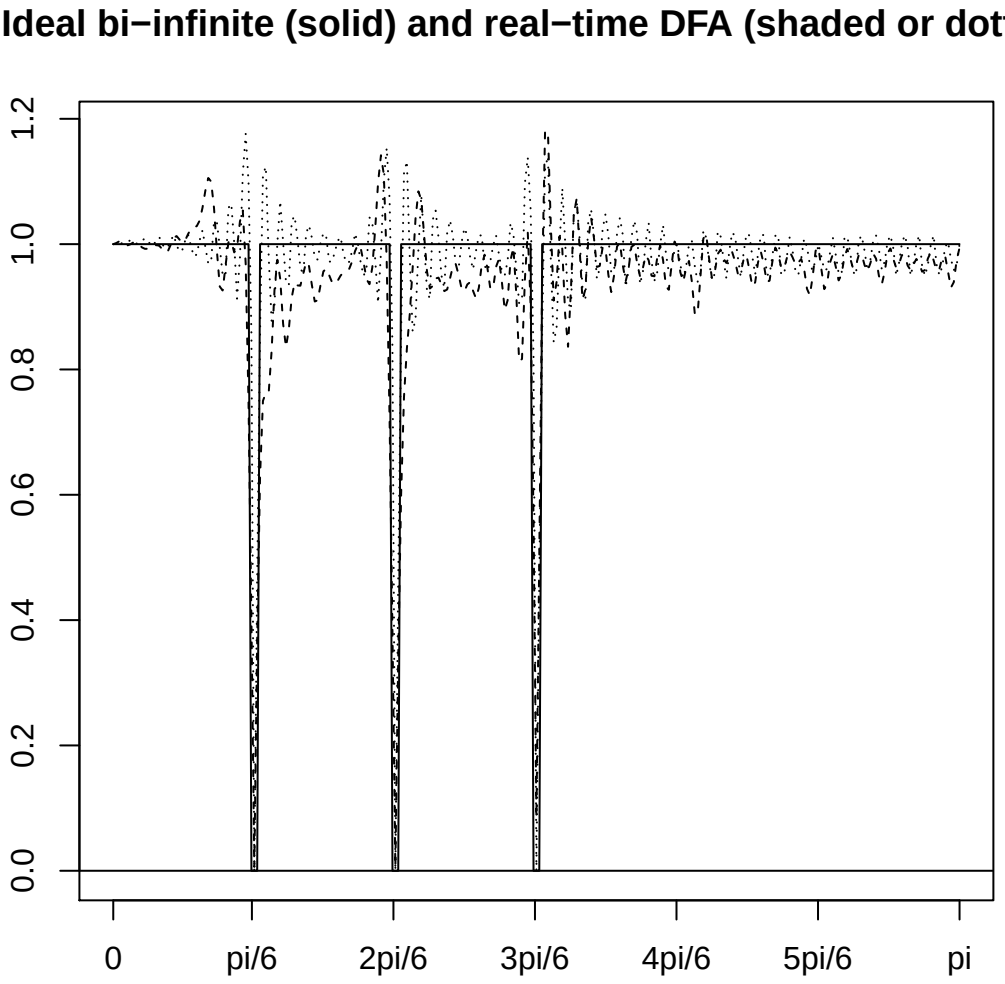

Figure 9: Amplitude functions of ideal bi-infinite target (solid) and finite one-sided DFA SA-filter based on the SEATS-spectrum (dotted) and on the periodogram (shaded).

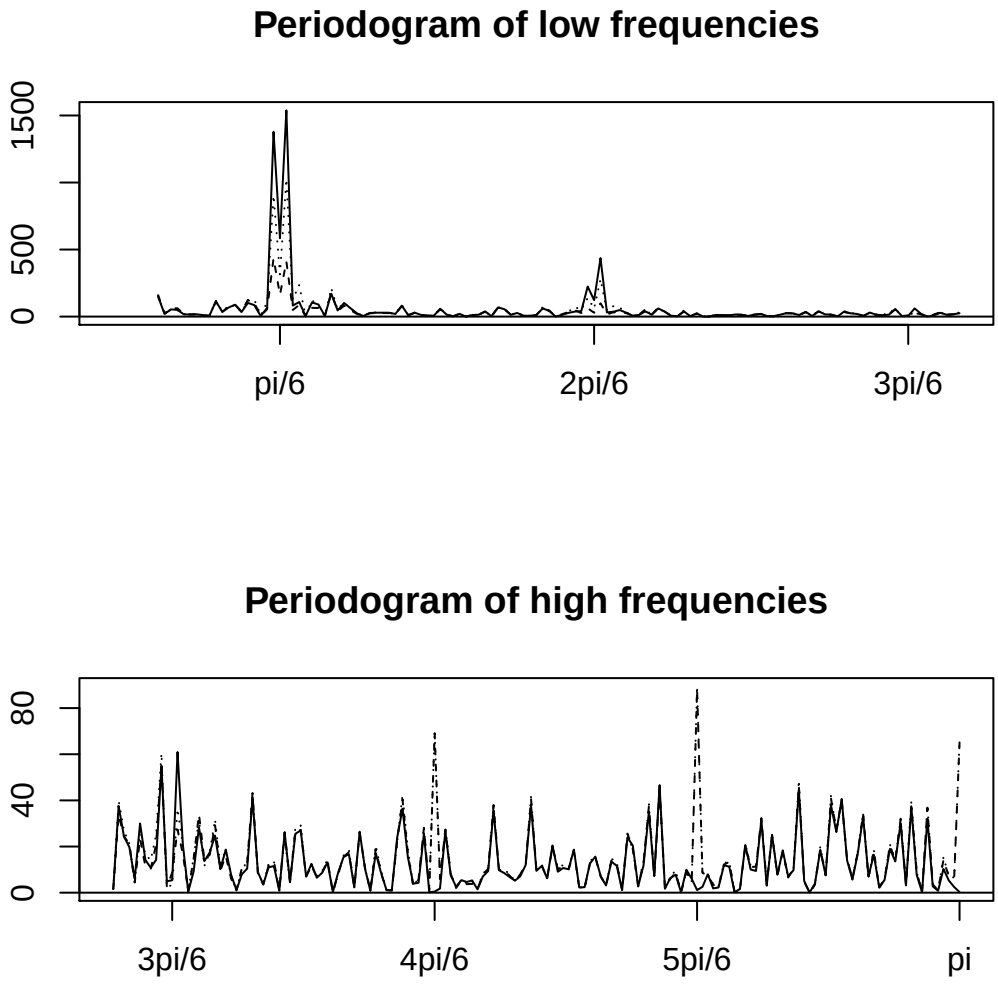

Figure 10: Periodograms of real-time MBA (solid), DFA based on the periodogram (shaded) and DFA based on the SEATS-spectrum (dotted) filter outputs splitted into dominant (top) and negligible (bottom) bands of seasonal frequencies.