International Journal of Forecasting 35 (2019) 1072–1084

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

The trilemma between accuracy, timeliness and smoothness in real-time signal extraction

Marc Wildi [a], Tucker S. McElroy [b,*]

[a] IDP, Zurich University of Applied Sciences, Rosenstrasse 8, 8401 Winterthur, Switzerland
[b] Research and Methodology Directorate, U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233- 9100, United States

ARTICLE INFO

Keywords:
Concurrent Filters
Frequency Domain
Pass-Band
Phase Delay
Trends

ABSTRACT

The evaluation of economic data and the monitoring of the economy is often concerned with an assessment of the mid- and long-term dynamics of time series (trend and/or cycle). Frequently, one is interested in the most recent estimate of a target signal, a so-called real-time estimate. Unfortunately, real-time signal extraction is a difficult estimation problem that involves linear combinations of possibly infinitely many multi-step ahead forecasts of a series. Here, we address the performances of real-time designs by proposing a generic direct filter approach. We decompose the ordinary mean squared error into accuracy, timeliness and smoothness error components, and we propose a new tradeoff between these competing terms, the so-called ATS-trilemma. This formalism enables us to derive a general class of optimization criteria that allow the user to address specific research priorities, in terms of the accuracy, timeliness and smoothness properties of the corresponding concurrent filter. We illustrate the new methods through simulations, and present an application to Indian industrial production data.

Published by Elsevier B.V. on behalf of International Institute of Forecasters.

1. Introduction

The evaluation of economic data and the monitoring of the economy is often concerned with an assessment of mid- and long-term dynamics of time series (trend and/or cycle). While this is acknowledged as an important task for developed nations, it is argued that the assessment of such dynamics for developing countries is even more crucial and challenging, given the greater volatility of such data. A broad range of techniques are available to the analyst: ARIMA-based approaches (e.g., TRAMO-SEATS; see Maravall & Caparello, 2004), state space methods (e.g., STAMP; see Koopman, Harvey, Doornik, & Shepherd, 2000), and classic filters (e.g., Hodrick-Prescott (HP) or Cristiano-Fitzgerald (CF); see Christiano & Fitzgerald, 2003, and Hodrick & Prescott, 1997). Frequently, one is interested in the most recent estimate of a target signal, which depends only on present and past data through a concurrent filter, and is referred to as a real-time estimate. Unfortunately, the performances of real-time estimates generally differ from those of historical estimates (based upon symmetric filters, also called smoothers) because the former cannot rely on future data, i.e., concurrent filters are asymmetric. We address the performances of real-time designs (although the method can be easily extended to smoothing) here by proposing a generalization of the direct filter approach (DFA) first introduced by Wildi (2005). The new idea contained in this paper is that real-time estimates can be constrained to have desirable properties — such as an increased smoothness and timeliness — by the direct design of the corresponding concurrent filter.

We impose these desirable properties on a filter by decomposing the signal extraction mean squared error (MSE) into components that correspond to the key properties of a filter. It is well known that any linear filter is characterized by its frequency response function (frf),

* Corresponding author.
E-mail addresses: marc.wildi@zhaw.ch (M. Wildi), tucker.s.mcelroy@census.gov (T.S. McElroy).

https://doi.org/10.1016/j.ijforecast.2019.03.008
0169-2070/Published by Elsevier B.V. on behalf of International Institute of Forecasters.

i.e., the discrete fourier transform (DFT) of its filter coefficients, and that the frf can be decomposed in terms of a gain function and a phase function. The former governs the smoothness of filter estimates, describing the so-called pass-band and stop-band, whereas the latter governs the timeliness of the filter, controlling the temporal advance or retardation of underlying harmonics. Here, we make a novel connection of these concepts to the signal extraction MSE. We decompose the ordinary MSE into accuracy, timeliness and smoothness error components, and propose a new tradeoff among these competing terms, the so-called ATS-trilemma. In a way, this is analogous to the bias–variance dilemma that arises from the classical decomposition of a parameter estimator's MSE, but it is a richer two-dimensional tradeoff among three competing terms which address signal extraction explicitly.

This formalism enables us to derive a general class of optimization criteria that allow the user to address specific research priorities, in terms of the accuracy, timeliness and smoothness properties of the corresponding concurrent filter. We call such a criterion a "customization". Although customization for real-time signal extraction can be achieved via other methodologies, we argue that our particular decomposition of the MSE offers the most direct and compelling connection between parameters and the corresponding characteristics of a concurrent filter. For example, in a model-based framework (e.g. TRAMO-SEATS or STAMP), one could adjust the model orders and/or model parameters in order to achieve modifications to smoothness (e.g. increase the integration order for the trend so as to obtain a smoother real-time trend), but the connection between such adjustments and the phase function of the concurrent filter that arises from such models is much harder to discern; moreover, knowing the phase function alone does not provide information about its impact on the MSE, and ultimately one is concerned about the signal extraction error. If one advocates using an asymmetric version of a popular nonparametric filter instead, such as the HP, CF, or ideal low-pass filter, there are few parameters available for adjusting the phase function, and the story is much the same as in the model-based framework. Thus, while we acknowledge that other methodologies provide a connection between parameters and vague notions of smoothness and phase, our contribution here is to make these connections explicit mathematically through the ATS decomposition. In fact, the ATS decomposition can be used to analyze and customize either a model-based concurrent filter or a nonparametric concurrent filter; we advocate working with a richer class of concurrent filters that allow for more flexibility in terms of gain and phase functions than is typically possible with model-based or nonparametric approaches.

If one were to weight the squared bias and variance by some convex combination, where equal weights correspond to the estimator's MSE, one could emphasize either aspect, depending upon user priorities. Such a customization results in a one-dimensional space — or curve — of criteria, with the classical MSE being only the central point. Similarly, real-time signal extraction customization results in a two-dimensional space — or triangle — of criteria, where the MSE lies in the center and each vertex corresponds to an exclusive emphasis on one of accuracy, timeliness, or smoothness. Thus, traditional approaches (such as those discussed above) can be replicated perfectly by DFA, and can also be customized. Interestingly, the two-dimensional tradeoff collapses to a bipolar dilemma in a classic one- or multi-step-ahead forecast perspective, as we demonstrate below (essentially because the notions of pass-band and stop-band are irrelevant in forecasting problems). The key point here is that the MSE can be decomposed into a summation of constituent error measures that correspond to useful quantities of interest. Just as with a classical estimator, where it is useful to examine both its squared bias and its variance — described heuristically as the accuracy and precision, respectively — it is useful here, in the context of signal extraction, to decompose the MSE into the three components of accuracy, timeliness, and smoothness. As in nonparametric density estimation, where altering the bandwidth can decrease the bias at the expense of a higher variance, or vice versa, an improvement in one of the ATS components here typically entails a deterioration in one or both of the other components.

The main concepts are introduced in Section 2. Section 3 then replicates and customizes a simple generic model-based filter, to demonstrate that DFA includes the classical approaches, while being more flexible. We propose classic time-domain performance measures, namely peak correlation (for timeliness) and curvature (mean square second-order differences for smoothness), and illustrate that the customized design can outperform model-based filters in both aspects simultaneously, out-of-sample. Due to space limitations, we restrict the exposition here to univariate approaches, though customization and the underlying trilemma readily extend to multivariate approaches, as is currently being investigated by the authors. Section 4 eschews the model-based framework and provides nonparametric results. Section 5 provides an application to the nowcasting of Indian industrial production data, with the finding of positive growth starting in 2015, based upon a customized filter. Section 6 summarizes our findings.

2. ATS-trilemma and customization

Here we introduce the mathematical concepts needed. Some of this material is provided by Wildi and McElroy (2016), but Section 2.3 here is novel, and exposits the main thesis of this paper.

2.1. Target

Let $\{x_t\}$ be our time series data. We assume that the target (signal) $\{y_t\}$ is specified by the output of a (possibly bi-infinite) filter applied to x_t:

$$y_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k}, \tag{1}$$

and we denote the frf of the filter by $\Gamma(\omega) = \sum_{k=-\infty}^{\infty} \gamma_k \exp(-ik\omega)$. This specification is completely general, such

that we could account for model-based targets (based on ARIMA or state space models), nonparametric filters (HP, CF, or Henderson, 1916), or "ideal" low-pass and band-pass filters such as

$$\Gamma(\omega;\eta_1,\eta_2) = 1_{[\eta_1,\eta_2]}(|\omega|), \tag{2}$$

where $0 \leq \eta_1 < \eta_2$ and with the convention that $\Gamma(\omega;\eta_1,\eta_2)$ is a low-pass if $\eta_1 = 0$. The ideal low-pass and band-pass filters are symmetric, with $\gamma_k = (\sin[\eta_2 k] - \sin[\eta_1 k])/(\pi k)$ for $k \geq 1$ and $\gamma_0 = (\eta_2 - \eta_1)/\pi$. Typically, signal extraction targets are symmetric ($\gamma_k = \gamma_{-k}$), but our definition in Eq. (1) allows for a general specification. For example, h-step-ahead forecasting can be obtained by setting $\gamma_k = 1_{\{k=-h\}}$. The resulting transfer function $\Gamma(\omega) = \exp(ih\omega)$ would be an anticipative all-pass filter.

2.2. Mean square paradigm

We seek a concurrent filter, called $\widehat{\Gamma}$ (because it is a concurrent approximation to Γ), of a given length $L > 0$. Let $\widehat{\Gamma}(B) = \sum_{k=0}^{L-1} b_k B^k$, expressed in terms of the backshift operator B. We seek filter coefficients $\{b_k\}$ such that the finite sample estimate

$$\widehat{y}_t := \sum_{k=0}^{L-1} b_k x_{t-k} \tag{3}$$

is as close as possible to y_t in mean square; i.e., we need to solve

$$\arg\min_{\mathbf{b}} E\left[(y_t - \widehat{y}_t)^2\right], \tag{4}$$

where $\mathbf{b} = [b_0, b_1, \ldots, b_{L-1}]'$. We could restrict $\widehat{\Gamma}(B)$ to filters that arise as the concurrent filter of a particular model (see Bell & Martin, 2004, for the formula of a concurrent filter as a function of signal and noise models), in which case there is a model parameter θ such that $\mathbf{b}$ is a function of θ. Typically, θ has a lower dimension than L, which may be fairly large (e.g., $L = 100$). Alternatively, $\widehat{\Gamma}(B)$ may be a concurrent version of a nonparametric filter, such as the Henderson trend filter. Here, too, $\mathbf{b}$ is a function of parameters θ that govern smoothing. In the Henderson case, θ is scalar and discrete, corresponding to the available (discrete) orders of Henderson filters. For the HP, θ corresponds to a scalar smoothing parameter. We also consider the case where $\widehat{\Gamma}(B)$ is not a function of an underlying parameter, but involves the full class of order L moving average filters; see Wildi and McElroy (2016).

For simplicity of exposition, we now assume that $\{x_t\}$ is a weakly stationary process with a continuous spectral density h (the DFT of the autocovariance function) such that

$$E\left[(y_t - \widehat{y}_t)^2\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\Gamma(\omega) - \widehat{\Gamma}(\omega)|^2 h(\omega)d\omega. \tag{5}$$

This follows from the fact that $y_t - \widehat{y}_t = (\Gamma(B) - \widehat{\Gamma}(B))x_t$, which has spectrum $|\Gamma(e^{-i\omega}) - \widehat{\Gamma}(e^{-i\omega})|^2 h(\omega)$; see Brockwell and Davis (1991). Generalizations to non-stationary integrated processes are proposed by Wildi (2005) and Wildi and McElroy (2016).[1] The replication of traditional model-based filter-designs is obtained by plugging the corresponding target signal $\Gamma(\omega)$ and the corresponding spectral density h (pseudo-spectral density in the case of integrated processes; see Bell & Hillmer, 1984, for a discussion) into Eq. (5).

2.3. Accuracy, timeliness and smoothness components

Consider the following identity:

$$\begin{aligned}
&|\Gamma(\omega) - \widehat{\Gamma}(\omega)|^2 \\
&= A(\omega)^2 + \widehat{A}(\omega)^2 - 2A(\omega)\widehat{A}(\omega)\cos\left(\Phi(\omega) - \widehat{\Phi}(\omega)\right) \\
&= (A(\omega) - \widehat{A}(\omega))^2 \\
&\quad + 4A(\omega)\widehat{A}(\omega)\sin^2\left(\frac{\Phi(\omega) - \widehat{\Phi}(\omega)}{2}\right),
\end{aligned} \tag{6}$$

where $A(\omega) = |\Gamma(\omega)|$ and $\widehat{A}(\omega) = |\widehat{\Gamma}(\omega)|$ are amplitude functions, and $\Phi(\omega) = Arg(\Gamma(\omega))$ and $\widehat{\Phi}(\omega) = Arg(\widehat{\Gamma}(\omega))$ are phase functions of the filters involved. In the case of typical signal extraction problems, where Γ is a symmetric filter, the phase of the target vanishes, i.e., $\Phi(\omega) = 0$. In the case of h-step-ahead forecasting, the phase becomes $\Phi(\omega) = h\omega$. We now plug Eq. (6) into Eq. (5) and obtain the following decomposition of the signal extraction MSE:

$$\begin{aligned}
\int_{-\pi}^{\pi} \left|\Gamma(\omega) - \widehat{\Gamma}(\omega)\right|^2 h(\omega)d\omega &= \int_{-\pi}^{\pi} (A(\omega) - \widehat{A}(\omega))^2 h(\omega)d\omega \\
&\quad + 4\int_{-\pi}^{\pi} A(\omega)\widehat{A}(\omega)\sin^2\left(\frac{\Phi(\omega) - \widehat{\Phi}(\omega)}{2}\right) h(\omega)d\omega.
\end{aligned}$$

In the case of signal extraction, we can specify the pass-bands and stop-bands of a filter as

$$\text{pass-band} = \{\omega | A(\omega) \geq 0.5\}$$
$$\text{stop-band} = [-\pi, \pi] \setminus \text{pass-band}.$$

The original MSE can then be decomposed additively into the following four terms:

$$\text{Accuracy} := \int_{\text{pass-band}} (A(\omega) - \widehat{A}(\omega))^2 h(\omega)d\omega \tag{7}$$

Timeliness

$$:= 4\int_{\text{pass-band}} A(\omega)\widehat{A}(\omega)\sin^2\left(\frac{\Phi(\omega) - \widehat{\Phi}(\omega)}{2}\right) h(\omega)d\omega \tag{8}$$

Smoothness

$$:= \int_{\text{stop-band}} (A(\omega) - \widehat{A}(\omega))^2 h(\omega)d\omega \tag{9}$$

Residual

$$:= 4\int_{\text{stop-band}} A(\omega)\widehat{A}(\omega)\sin^2\left(\frac{\Phi(\omega) - \widehat{\Phi}(\omega)}{2}\right) h(\omega)d\omega. \tag{10}$$

Accuracy measures the contribution to the MSE when the phase (time-shift) and the noise suppression in the stop-band are ignored. This would correspond to the performance of a symmetric filter (no time shift) with perfect noise suppression, i.e., $\widehat{A} = 0$ in the stop-band, and with the same amplitude as the concurrent filter considered in

[1] Under suitable filter restrictions, the filter error $y_t - \widehat{y}_t$ is weakly stationary even if $\{x_t\}$ is integrated.

the pass-band. A corresponding symmetric filter could be constructed easily by an inverse Fourier transformation. *Smoothness* measures the MSE contribution that can be attributed to the leakage of the concurrent filter in the stop-band, corresponding to undesirable high-frequency noise. This quantity is linked to the well-known time-domain "curvature" measure described in Section 3.2. *Timeliness* measures the MSE contribution generated by the time-shift. It is linked to the time-domain "peak correlation" concept described in Section 3.2. Finally, the *residual* is the part of the MSE which cannot be attributed to any of the error components above. Since the product $A(\omega)\widehat{A}(\omega)$ is small in the stop-band (possibly vanishing, as in the case of the ideal low-pass), the residual is generally negligible. From a slightly different perspective, user priorities are rarely concerned about the time-shift properties of components in the stop-band, which ought to be damped or eliminated anyway. Thus, for the sake of simplicity, we focus henceforth on the ideal low-pass target, where the residual vanishes completely.

2.4. Signal extraction ATS-trilemma and customization

The best MSE concurrent filter provides a benchmark; however, it can be beaten by other filters according to the facets of accuracy, smoothness, or timeliness. It is not possible for another filter to improve upon the benchmark according to all three of these facets, because otherwise such a filter would have a better overall MSE (omitting the residual in this discussion), but we may possibly improve two aspects simultaneously, at the expense of the third. For example, it can be shown that we can always improve the accuracy and smoothness, at the expense of an increased phase delay. The MSE criterion balances these terms, being a straight summation of Eqs. (7) and (8), and (9); weighting these terms unevenly will enforce different emphases, giving rise to quite different concurrent filters.

More formally, the MSE can be generalized easily to provide a customized measure by assigning weights to the terms of its ATS decomposition:

$$\mathcal{M}(\vartheta_1, \vartheta_2) = \vartheta_1 \,\text{Timeliness} + \vartheta_2 \,\text{Smoothness} + (1 - \vartheta_1 - \vartheta_2)\,\text{Accuracy}. \tag{11}$$

The parameters $\vartheta_1, \vartheta_2 \in [0, 1]$ (such that $\vartheta_1 + \vartheta_2 \leq 1$) allow one to assign priorities to single or pairwise combinations of MSE error terms. The resulting two-dimensional tradeoff is called the ATS-trilemma, and the optimization paradigm in Eq. (11) is called a customized criterion. Plotting $\mathcal{M}(\vartheta_1, \vartheta_2)$ on the ϑ_1- and ϑ_2-axes yields a surface that is defined over a triangle with vertices (0, 0), (1, 0), and (0, 1), which we call the "customization triangle". The user is free to navigate on the customization triangle according to specific research priorities. The ordinary MSE criterion (referred to as the "default" henceforth) is obtained by setting $\vartheta_1 = \vartheta_2 = 1/3$, whereas the edges correspond to the complete absence of one of the A, T and S terms, and the vertices emphasize one component to the exclusion of both others.[2]

The criterion in Eq. (11) is also impacted by the choice of pass-band; altering the pass-band from the definition $\{\omega | A(\omega) \geq 0.5\}$ changes the corresponding values of the accuracy, timeliness, and smoothness terms, which in turn affects the resulting customized filter. For instance, allowing the pass-band to encompass a greater range requires our customized filter to be timely (say, if $\vartheta_1 > 0$) across a greater range of frequencies.

The schematic criterion in Eq. (11) could be extended usefully by introducing non-negative weighting functions $W_1(\omega)$ and $W_2(\omega)$ to the timeliness and smoothness terms:

$$\text{Timeliness} = 4 \int_{\text{pass-band}} A(\omega)\widehat{A}(\omega) \sin^2\left(\frac{\Phi(\omega) - \widehat{\Phi}(\omega)}{2}\right) \times W_1(\omega) h(\omega) d\omega \tag{12}$$

$$\text{Smoothness} = \int_{\text{stop-band}} (A(\omega) - \widehat{A}(\omega))^2 W_2(\omega) h(\omega) d\omega. \tag{13}$$

These weighting functions can be viewed as extensions of the constant weights ϑ_1 and ϑ_2 that appear in Eq. (11), which weight timeliness and smoothness, respectively. For example, we may wish to de-emphasize higher frequencies in the smoothness term, in order to avoid the incidence of false turning points in the real-time signal, and this could be accomplished by letting $W_2(\omega)$ take larger values for ω in the stop-band. While there is no best choice for such weighting functions, their design should correspond to the priorities of each user; we provide an example in the empirical section below.

2.5. Forecast AT-dilemma

The above trilemma was obtained by assuming that the target filter Γ discriminates the components into pass- and stop-bands. In contrast, h-step-ahead forecasting is concerned with the approximation of an anticipative all-pass filter. Since the stop-band is non-existent, the smoothness and residual error components vanish and the above trilemma collapses into a forecasting dilemma:

$$\mathcal{M}(\vartheta_1) = \vartheta_1 \,\text{Timeliness} + (1 - \vartheta_1)\,\text{Accuracy}.$$

This function $\mathcal{M}(\vartheta_1)$ could be plotted against $\vartheta_1 \in [0, 1]$, yielding a curve. While this paradigm is sufficient to address all-pass filtering problems, such as multi-step-ahead forecasting, it cannot emphasize smoothness, and hence is not sufficient for more general real-time estimation problems, for which Eq. (11) is preferred.

[2] Fixing the length L of the concurrent filters means that each of the A, T and S terms will be finite, ensuring that no numerical problems result when customizing on the triangle's edges.

2.6. Spectrum, signal and customization interfaces

The proposed method can replicate and customize both traditional (ARIMA or state space) model-based approaches and classic filter designs (HP, CF, Henderson) by plugging the corresponding spectral densities and the target signals into Eq. (11). We could combine these in any order; for example, we could target a HP filter by supplying a model-based spectral density. Moreover, we could use alternative spectral estimates (e.g., nonparametric) or targets: the resulting spectrum- and signal-interfaces allow for a flexible implementation of general forecasting and signal extraction problems. Examples of hybrid real-time trend extraction and seasonal adjustment problems are provided by Wildi and McElroy (2016).

The customization interface in Eq. (11) allows one to emphasize particular filter characteristics, so as to align with one's research priorities. For instance, one can set the pass-band and stop-band according to the frequencies that are of greatest interest for the application. One can also use the criterion in Eq. (11) to obtain customized filters subject to some constraint (this would be accomplished by minimizing $\mathcal{M}(\vartheta_1, \vartheta_2)$ subject to a hard constraint on the filter coefficients $\widehat{\Gamma}$); e.g., constraining timeliness to equal some specified value, and searching for the real-time filter that has the best possible accuracy and smoothness given this condition.

3. Replicating and customizing a model-based approach

3.1. Empirical design

We propose a simple simulation framework here, based on an ideal low-pass target with cutoff $\pi/12$ given by $\Gamma(\omega; 0, \pi/12)$ (which implies that $\gamma_k = \frac{\sin(k\pi/12)}{k\pi}$ for $k \neq 0$ and $\gamma_0 = \frac{1}{12}$) and an AR(1) data generating process (DGP) $\{x_t\}$ that satisfies

$$x_t = ax_{t-1} + \epsilon_t, \tag{14}$$

where $\{\epsilon_t\}$ is a white noise process. Our framework is pertinent to real-time economic indicators: log-returns of macro-data[3] can be fitted more or less satisfactorily by benchmark AR(1) models, and users are typically interested in inferring business cycle or trend dynamics from such a design. Our framework is hybrid in the sense that we target a non model-based signal (the ideal trend) based upon a model of the DGP. We deliberately oversimplify our artificial framework in order to highlight the salient features of the ATS-trilemma.

As was mentioned above, we can introduce weighting functions to the timeliness (Eq. (12)) and smoothness (Eq. (13)) terms. Here, letting $\pi/12$ be the cutoff of the target signal, we adopt the weighting functions

$$\left.\begin{array}{rcl} W_1(\omega;\lambda) & = & 1+\lambda \\ W_2(\omega;\eta) & = & (1+|\omega|-\pi/12)^{\eta} \end{array}\right\} \tag{15}$$

[3] Industrial production, employment or income, as proposed and used by the NBER.

and utilize the criterion in Eq. (11) with the basic definition of the accuracy term in Eq. (7). For $\lambda = \eta = 0$, the equally-weighted criterion (i.e., the MSE benchmark) is obtained; $\lambda > 0$ emphasizes timeliness uniformly across frequencies (and is equivalent to increasing ϑ_1), whereas adopting $\eta > 0$ highlights smoothness by further penalizing amplitude discrepancies in the stop-band. We can balance the impacts of timeliness and smoothness by taking moderate values of both λ and η; we have found by experimentation that $\lambda = 30$ and $\eta = 0.5$ are adequate for this purpose.

3.2. Performance measures

We rely on the following time-domain measures for assessing the out-of-sample performances of our concurrent filter designs:

$$\text{Curvature} := \frac{E\left[\left((1-B)^2\widehat{y}_t\right)^2\right]}{\text{var}(\widehat{y}_t)} \tag{16}$$

$$\text{Peak correlation} := \text{Arg}\left(\max_j(cor(y_t, \widehat{y}_{t+j}))\right) \tag{17}$$

$$\text{Sample MSE} := E\left((y_t - \widehat{y}_t)^2\right). \tag{18}$$

Mean square second-order differences emphasize the geometric curvature of a time series; a smaller curvature leads to a smoother appearance of a filtered series (e.g., having less noisy ripples). The proposed measure is normalized in order to immunize against scaling effects. Note that we are interested primarily not in the magnitude of the correlations in Eq. (17), but in the integer j_0 at which the correlation between the target y_t and the shifted (real-time) estimate $\widehat{y}_{t+j_0}$ is maximized. The latter is called leading, coincident or lagging based on whether j_0 is negative, zero or positive, respectively. The sample MSE is computed in order to assess the overall loss in mean square performances that is entailed by customization.

Since expectations are unknown, we will report sample distributions (box plots) of the above performance measures, distinguishing between in- and out-of-sample periods. In particular, we will benchmark four empirical filters — three customized and one default (i.e., taking $\lambda = \eta = 0$ and $\vartheta_1 = \vartheta_2 = 1/3$ in Eq. (11)) — against the best possible filter (assuming a knowledge of the true AR(1) DGP).

3.3. A generic model-based approach

Given a sample $x_1, \ldots, x_T$ of $\{x_t\}$, the generic model-based approach to obtaining a concurrent filter consists of forecasting and backcasting missing data in the target specification. Letting $t = T$ in Eq. (1), we obtain

$$y_T = \sum_{k=-\infty}^{-1} \gamma_k x_{T-k} + \sum_{k=0}^{T-1} \gamma_k x_{T-k} + \sum_{k=T}^{\infty} \gamma_k x_{T-k},$$

which partitions the target into future observations, available data, and past observations. The future and past sections are unobserved, so the corresponding x_{T-k} must be replaced by a projection $\widehat{x}_{T-k}$, which is a forecast if

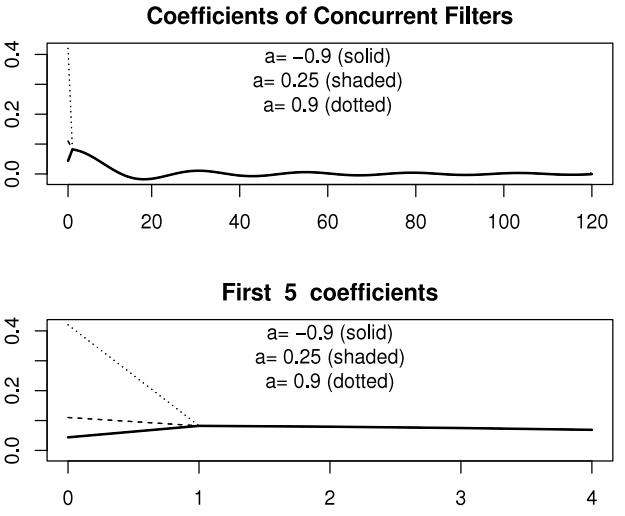

Fig. 1. Coefficients of concurrent filters for three different AR(1) processes: the model-based coefficients and the estimates based on the DFA MSE are indistinguishable.

$k < 0$ and a backcast if $k \geq T$. Assuming that the AR(1) model is correct, we obtain $\widehat{x}_{T-k} = a^{-k}x_T$ for forecasts and $\widehat{x}_{T-k} = a^{k-T+1}x_1$ for backcasts. Substituting these into the expression for y_T, we obtain

$$\widehat{y}_T = \sum_{k=-\infty}^{-1} \gamma_k a^{-k} x_T + \sum_{k=0}^{T-1} \gamma_k x_{T-k} + \sum_{k=T}^{\infty} \gamma_k a^{k-T+1} x_1$$
$$= \left(\sum_{k=0}^{\infty} \gamma_{-k} a^{-k} \right) x_T + \sum_{k=1}^{T-2} \gamma_k x_{T-k} + \left(\sum_{k=0}^{\infty} \gamma_{T-1+k} a^{k} \right) x_1.$$

Comparing this expression with Eq. (3) and setting $L = T$, we see that

$$b_0 = \sum_{k=0}^{\infty} \gamma_{-k} a^{-k}, \quad b_{T-1} = \sum_{k=0}^{\infty} \gamma_{T-1+k} a^{k},$$

and $b_k = \gamma_k$ for $1 \leq k \leq T-2$. Clearly, $b_{T-1} \approx 0$ for large values of T. This calculation can be generalized to ARMA processes (Wildi & McElroy, 2016). The formula explicitly emphasizes the forecast paradigm underlying classic model-based approaches.

We now compare the above filter coefficients b_j with those obtained using Eq. (11), assuming $\lambda = \eta = 0$ (continuous integrals are substituted by discrete sums). Fig. 1 illustrates the outcomes for three different DGPs based on $a = -0.9, 0.25, 0.9$ in Eq. (14). The middle choice $a = 0.25$ corresponds to an AR(1) model fitted to first differences of the (log-transformed) monthly industrial production index, estimated on data from January 1990 to February 2016 (the model passes ordinary diagnostic checks). The high value $a = 0.9$ corresponds to a process with a greater frequency content at the low frequencies, whereas $a = -0.9$ corresponds to a spectral density that peaks at frequency π, and has low values in the pass-band. These cases, which are included for illustrative purposes, represent different challenges for the ATS criterion, as the A and T terms will be large and S will be small when $a = 0.9$ (and vice versa when $a = -0.9$). We set $L = T = 120$ so that the filter and sample lengths correspond to ten years of monthly data. The DFA and model-based coefficients virtually overlap for all three DGPs, which confirms the ability of DFA to replicate the model-based approach. The approximation can be improved arbitrarily by selecting a denser frequency-grid when approximating the continuous integral in Eq. (5) by a discrete sum.

3.4. Customizing the model-based approach: known DGP

We first assume a knowledge of the true DGP, although this unrealistic assumption is relaxed in Section 3.5. This is accomplished by setting $h(\omega) = |1 - ae^{-i\omega}|^{-2}$ in Eqs. (7)–(9), and (10). Besides our benchmark filter (solid line in the following figures), corresponding to $\lambda = \eta = 0$ in Eq. (15), we propose to compute and analyze three customized designs: a "balanced" filter (dotted line, corresponding to $\lambda = 30$ and $\eta = 0.5$) that will outperform the benchmark filter in terms of the timeliness *and* smoothness components, and two unbalanced designs that emphasize more heavily the facets of either timeliness (large dots, corresponding to $\lambda = 500$ and $\eta = 0.3$) or smoothness (solid bold, corresponding to $\lambda = 30$ and $\eta = 1$). The empirical distributions of the performance measures are based on simulated time series of length $T = 1000$ (100 replications).

The amplitude and time-shift functions of all filters for $a = 0.25$ are plotted in Fig. 2 (the corresponding Figures A.1 and A.2 for the DGPs $a = -0.9$ and $a = 0.9$ are to be found in the supplement). Increasing λ tends to decrease the time-shift in the pass-band, as desired. For example, the timeliness filter (large dots) is a virtually zero-phase concurrent filter. Augmenting η means that the amplitude function in the stop-band is pulled more strongly towards the target (which is zero). Augmenting λ and η simultaneously implies a weaker control of the amplitude function in the pass-band by the accuracy term.

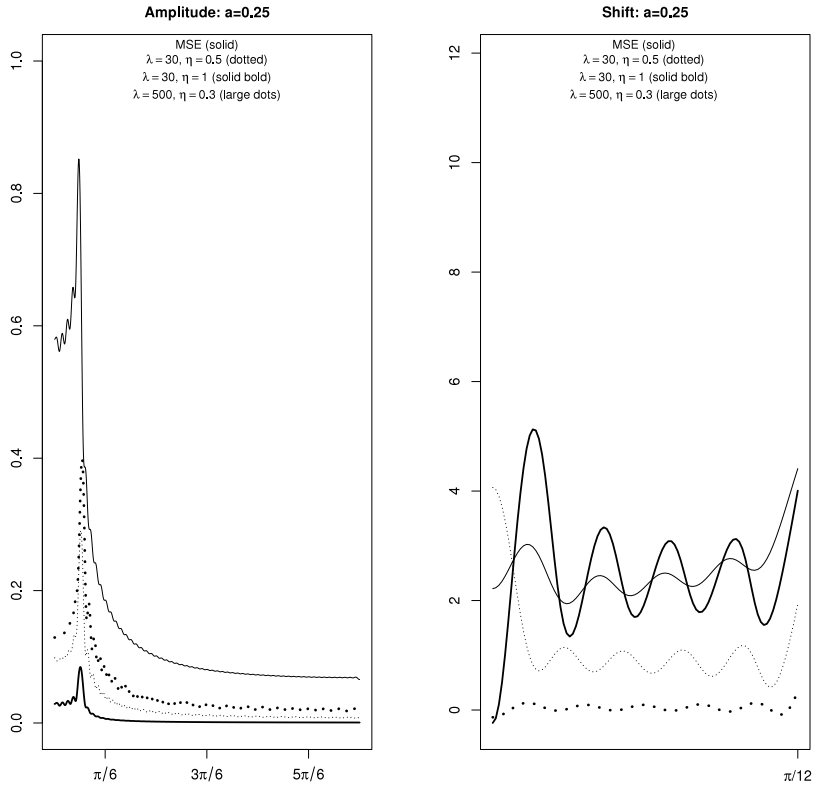

Fig. 2. Amplitude and time-shift plots for MBA and customized filters, for AR(1) DGP with $a = 0.25$. The time-shift is restricted to the pass-band of the filters.

Table 1
Total MSE and ATS error components for the MBA default and customized filters, for an AR(1) DGP with $a = 0.25$.

	Accuracy	Timeliness	Smoothness	Residual	Total MSE
MBA-MSE	0.021153	0.018681	0.024016	0.000000	0.063850
$\lambda = 30,\ \eta = 0.5$	0.114323	0.000512	0.002403	0.000000	0.117237
$\lambda = 30,\ \eta = 1$	0.136593	0.000996	0.000151	0.000000	0.137740
$\lambda = 500,\ \eta = 0.3$	0.104476	0.000008	0.006450	0.000000	0.110934

The empirical distributions of the curvature and peak correlation[4] in Fig. 3 confirm the intended effects, namely that both dimensions can be addressed either individually (large dots, solid bold) or simultaneously (dotted) by suitable customizations (the other two processes can be seen in Figures A.3 and A.4 in the supplement). A stronger customization is accompanied by poorer MSE performances, as expected.[5]

Table 1 reports the true MSEs and the ATS-decompositions of the various designs for $a = 0.25$ (Tables A.1 and A.2 , corresponding to $a = -0.9$ and $a = 0.9$, are to be found in the supplement). As expected, the timeliness and smoothness components are affected by customization.[6]

To conclude, we plot filter outputs for a particular realization (the first of the sample) in Fig. 4; the other two processes are to be seen in Figures A.5 and A.6 in the supplement. A symmetric filter of length 121 is used for reference. All series are standardized for ease of visual inspection. Series with smaller peak correlations tend to

[4] The distribution of the peak correlation is concentrated because the time series are long (1000 observations).

[5] Part of the loss in MSE performances is due to the zero-shrinkage of the customized filter coefficients, and could be avoided by a straightforward re-scaling of the filter outputs (these results are not shown here).

[6] The shrinkage effect of heavily customized filters (solid bold) distorts these numbers to some extent, such that a direct comparison across filters is less straightforward than with the scale-invariant performance measures in Fig. 3, A.3 and A.4.

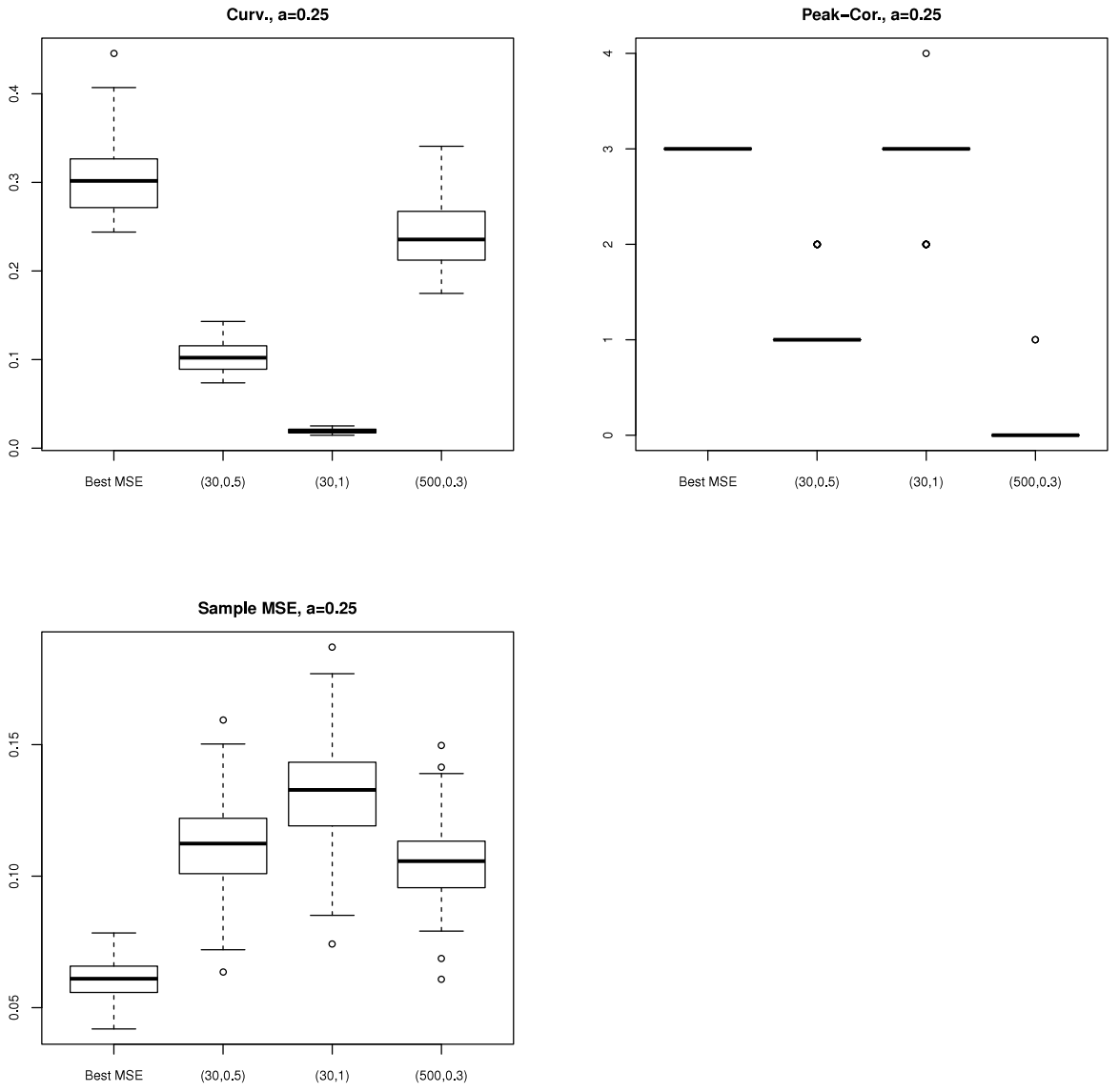

Fig. 3. Empirical distributions of the curvature, peak correlation and sample MSE, with $a = 0.25$.

be shifted leftwards with respect to the benchmark filter outputs (solid), while series with smaller curvatures appear to be smoother, as expected.

3.5. Customizing the model-based approach: empirical AR(1) spectrum

The results thus far rely on the unrealistic assumption that the true DGP is known. We now analyze the empirical spectral estimates based on estimated AR(1) processes, by plugging the estimated autoregressive spectrum in for $h(\omega)$. All estimates rely on samples of length $T = 120$. Out-of-sample performances are computed on samples of length 1000, and empirical distributions of performances rely on 100 replications of this basic setting for each of the above AR(1) processes. We avoid identification issues here by assuming that the model is known. Thus, an empirical AR(1) spectrum is obtained from the estimated AR(1) coefficient for each realization. In- and out-of-sample empirical distributions of the curvature, peak correlation and sample MSE are plotted in Fig. 5 for the DGP $a = 0.25$. The other two processes are to be found in the supplement, in Figures B.1 and B.2 . The out-of-sample distributions appear to be concentrated more tightly because the sample lengths are longer (1000 versus 120 in-sample). The balanced design ($\lambda = 30$, $\eta = 0.5$) clearly outperforms the benchmark filter in terms of smaller peak correlation and curvature measures, for all three processes, though at a cost in MSE performance.

4. Nonparametric spectrum

In contrast to the previous section, here we ignore any a priori knowledge about the true DGP and use a

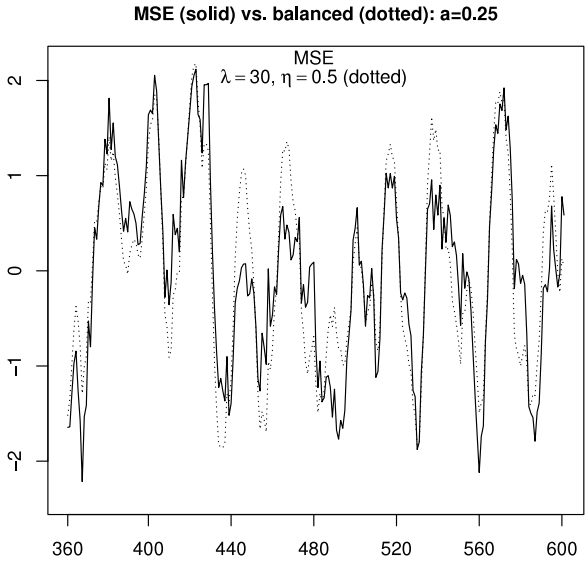

Fig. 4. Filter outputs from data simulated from an AR(1) DGP ($a = 0.25$), comparing historical symmetric (solid) estimates with real-time estimates (dotted), using both the benchmark and the customized designs.

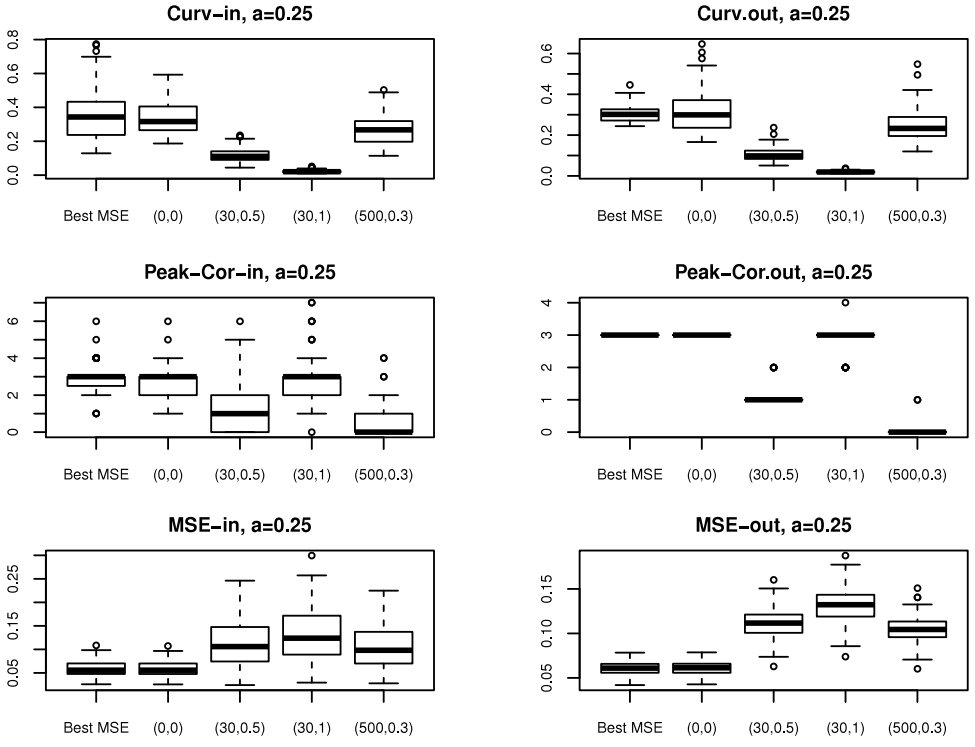

Fig. 5. Empirical distributions of the curvature, peak correlation and sample MSE based on an estimated AR(1) spectrum, with both in-sample (left plots) and out-of-sample (right plots) results given for $a = 0.25$.

nonparametric spectral estimate, namely the raw periodogram. We rely on the same empirical setting as in the previous section but set $L = 24$, and the periodogram is computed on the discrete frequency grid $\omega_k = k2\pi/T$, $k = -T/2, \ldots, T/2$, where $T = 120$ (ten years of monthly data). The choice of the filter length reflects the maximal duration of components in the stop-band of the target filter (two years). In addition to illustrating flexibility, here we also assess overfitting in a richly parametrized framework.

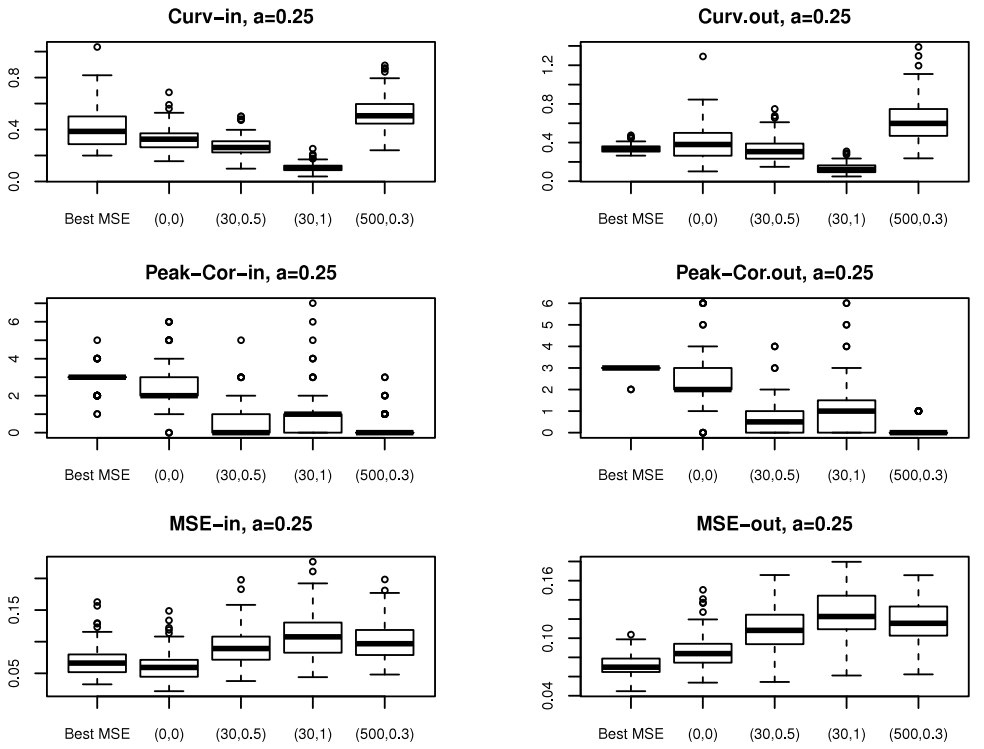

Fig. 6. Empirical distributions of the curvature, peak correlation and MSE based on periodograms: in-sample (left plots) and out-of-sample (right plots) for $a = 0.25$.

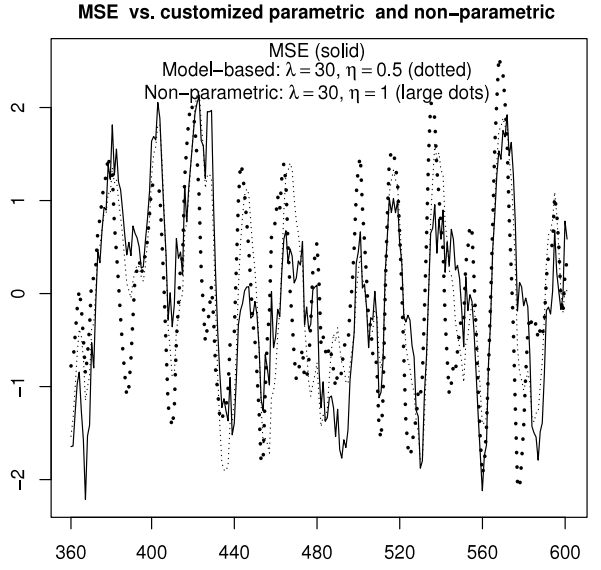

Fig. 7. Outputs of the best benchmark (solid) and of customized model-based (dotted) and customized nonparametric (large dots) filters, for a realization of an AR(1) process with $a = 0.25$.

The in- and out-of-sample empirical distributions of the curvature, peak correlation and sample MSE are plotted in Fig. 6 (see also Figures C.1 and C.2 in the supplement for the other two processes). The out-of-sample MSE performances are only slightly worse than the previous AR(1) spectrum. The empirical distributions of the peak correlation appear to be more dispersed, while the customization effect appears to be shifted. The peak correlations are smaller and the curvatures are larger than those obtained in the previous exercise that utilized an estimated parametric spectrum. The latter effect is due to the fact that the pass-band is short (only six discrete frequency ordinates, including frequency zero) and the number of freely-determined parameters $L = 24$ is comparatively large; thus, it is easier to emphasize time-shift properties in the pass-band (timeliness) than amplitude characteristics in the stop-band (smoothness). Despite the problem of overfitting, we observe that suitably customized filters ($\lambda = 30$, $\eta = 1$) still outperform the benchmark design in the targeted dimensions, according to out-of-sample criteria. Also, the out-of-sample and in-sample performances are remarkably consistent, given the fact that $L = 24$ coefficients are estimated in samples of length $T = 120$.

We illustrate the above results visually by briefly comparing the outputs of the best benchmark filter (assuming a knowledge of the DGP), the balanced customized model-based filter (Section 3.5) and the nonparametric customized estimate (previous graphs) for a particular realization (the first of the sample); see Fig. 7. The customized filter outputs lie to the left of the benchmark filter, and their appearance is smoother, in the sense that noisy high-frequency ripples are damped more effectively.

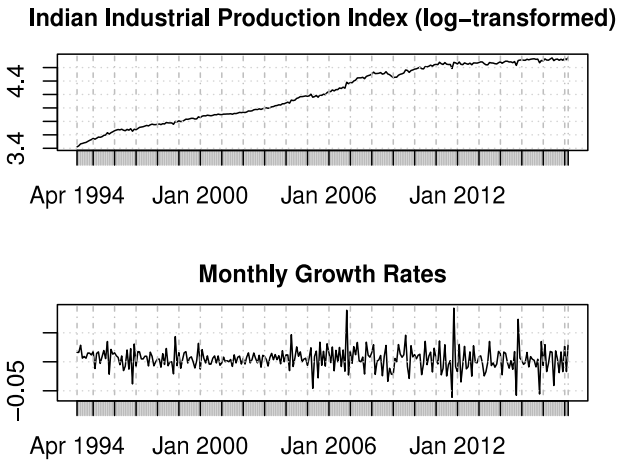

Fig. 8. Indian industrial production index (monthly and seasonally adjusted).

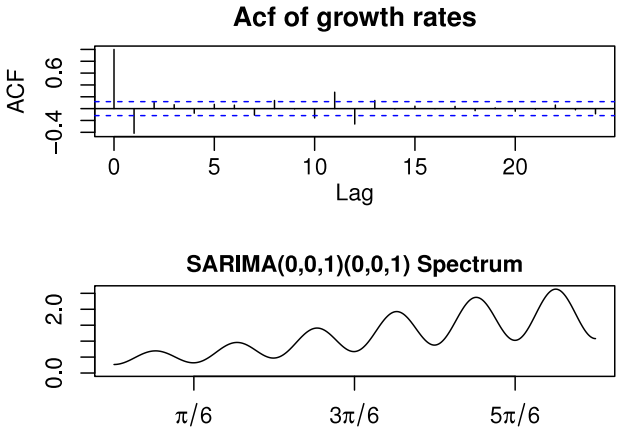

Fig. 9. Indian industrial production index: acf of growth rates (top panel) and the model-based spectrum (bottom panel).

5. Nowcasting secular growth in Indian industrial production

We analyze the production of total industry in India (https://fred.stlouisfed.org/series/INDPROINDMISMEI) from April 1994 to May 2017, a monthly seasonally-adjusted series that is shown in the top panel of Fig. 8. It is clear that the secular growth of the series evolves over time. We propose to analyze this phenomenon by applying suitably parametrized concurrent filters to the monthly growth rates displayed in the bottom panel of Fig. 8.

5.1. Model-based approach

An inspection of the autocorrelation plot in the top panel of Fig. 9 reveals a significant negative serial dependence in the growth rates at lags one and twelve, which is compatible with a model-based seasonal adjustment of the series (see McElroy, 2012, for discussion). A SARIMA(0,0,1)(0,0,1) model provides an adequate description of the monthly growth rates. The resulting model-based spectral density

$$S(\omega_k) = |(1 - 0.34\exp(-i\omega_k))(1 - 0.22\exp(-i12\omega_k))|^2$$

is shown in the bottom panel of Fig. 9. We capture the long-term (secular) trend growth by applying an ideal low-pass filter with cutoff $\pi/72$, corresponding to a mean cycle duration of $2 \times 72/12 = 12$ years. The amplitude function of the corresponding concurrent MSE filter is shown in Fig. 10; note that the amplitude peaks at frequency 0.028, corresponding to a mean cycle duration of 9.2 years. This unusually tight specification allows the secular trend to be extracted by smoothing out the business cycle.

For sake of comparison, we benchmark these performances against the simple exponential smoothing approach (Hyndman, Koehler, Ord, & Snyder, 2008), which amounts to generating forecasts according to an ARIMA(0,1,1) model. (Double exponential smoothing corresponds to an ARIMA(0,2,2) model, and is unlikely to produce reasonable forecasts given that the data have already been differenced to stationarity.) The moving average parameter is obtained by fitting the model, and the estimate is close to -1 because the data are stationary.

(We also experimented with other fixed values of the moving average parameter, such as 0.8, 0.5, and 0.2; lower values correspond to forecasts that depend highly on the most current observation, and higher values generate forecasts that depend highly on the distant past. However, the real-time performance was worse, so we present the results using an estimated smoothing parameter.) Extending the sample by forecasting in this manner allows the given target filter to be applied, resulting in a real-time signal extraction estimate generated by exponential smoothing.

5.2. Customization

When attempting to capture the long-term movements of industrial production, we require a real-time trend that is both smooth (so that we avoid false turning points) and timely. This objective is particularly important towards the end of the sample, where some interesting behavior occurs. To this end, we select $\lambda = 60$ and $\eta = 1$ in our design of a customized filter, thereby emphasizing timeliness and smoothness in the ATS-trilemma. Note that we rely on the above model-based spectrum for deriving both the MSE and the customized filters, of which the corresponding outputs (rescaled for comparability) are displayed in Fig. 11.

As expected, the customized filter output is smoother and lies to the left of the classic model-based design. The classic MSE filter has the broad swings (of period 12 years) that we hope that a low-pass filter would provide, but is accompanied by many smaller movements, or noise, which can lead locally to the detection of spurious turning points. The customized real-time signal has all such high frequency noise eliminated (this is the contribution of the smoothness term), while also preserving the chief low frequency swings (due to the presence of the accuracy term in our criterion). Moreover, the customized real-time signal overall is more timely than the classic signal, and leads by roughly a year in the 2001 nadir.

In contrast, the signal from exponential smoothing is similar to the classic real-time estimate qualitatively in terms of both smoothness and accuracy, but resembles the customized filter in terms of timeliness. Put another way, the customized real-time signal is a much smoother version of the signal from exponential smoothing. More interestingly, the customized filter reveals the presence of an up-turn starting in 2015, which indicates a possible relaxation of the zero-growth period towards the end of the sample. This feature of the series is missed by both the classic MSE filter and the simple exponential smoothing filter, with the outputs of the latter being subject to excessive lag and noise.

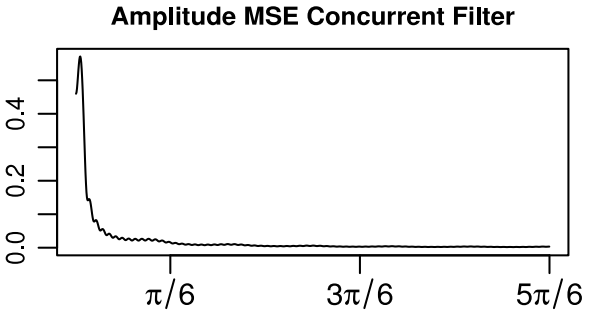

Fig. 10. Amplitude function of concurrent MSE filter.

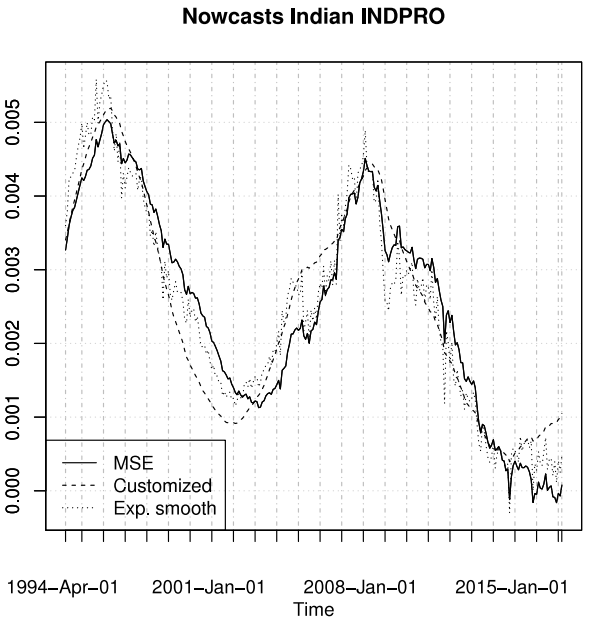

Fig. 11. Concurrent filter outputs: classic MSE (solid) vs. simple exponential smoothing (dotted) vs. customized (shaded).

6. Summary

Real-time signal extraction is a difficult prospective estimation problem that involves linear combinations of (possibly infinitely many) multi-step-ahead forecasts of a series. Addressing this problem directly with the DFA allows one to expand the forecast AT-dilemma into a richer ATS-trilemma. The resulting customized optimization criterion generalizes the classical MSE paradigm by allowing the user to utilize the customization triangle according to specific research priorities. In particular, the smoothness (curvature) and timeliness (peak correlation) performances of a concurrent filter can be improved simultaneously, at the expense of the accuracy and total MSE.

The DFA and the ATS-trilemma are generic concepts, in the sense that they allow for arbitrary target signals or spectral estimates. In particular, classical ARIMA-based approaches can be replicated exactly by supplying the corresponding entries. Then, once replicated, a particular approach can be customized to accommodate particular research priorities. Our numerical examples confirm that the best MSE filter, assuming a knowledge of the true DGP, can be outperformed predictably in terms of the curvature and peak correlation out-of-sample by suitably customized parametric (model-based) or nonparametric (periodogram) spectral estimates. When relying

on the periodogram, overfitting may affect MSE performances as well as the customization effect, especially if the pass-band is short.

These techniques can be applied to the real-time trend estimation, or nowcasting, of macroeconomic data, such as the gross domestic product or industrial production. Whereas larger economies, and in particular the economies of developed nations, tend to be stable (apart from in times of crisis), and hence more easily forecastable, the economies of developing countries can be subject to many vicissitudes of policy or political instability. As a result, the detection of turning points in a timely fashion is extremely important, and our analysis of Indian Industrial Production illustrates the utility of filter customization, where we were able to improve the timeliness and smoothness of real-time estimates simultaneously, while preserving the overall signal features.

Acknowledgments

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.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2019.03.008.

References

Bell, W., & Hillmer, S. (1984). Issues involved with seasonal adjustment of economic time series. *Journal of Business & Economic Statistics*, *2*, 291–320.

Bell, W., & 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.

Brockwell, P., & Davis, R. (1991). *Time series: Theory and methods*. New York: Springer.

Christiano, L., & Fitzgerald, T. (2003). The band pass filter. *International Economic Review*, *44*, 435–465.

Henderson, R. (1916). Note on graduation by adjusted average. *Transactions of the Actuarial Society of America*, *17*, 43–48.

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

Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer Science & Business Media.

Koopman, S., Harvey, A., Doornik, J., & Shepherd, N. (2000). *Stamp 6.0: Structural time series analyser, modeller, and predictor*. London: Timberlake Consultants.

Maravall, A., & Caparello, G. Program TSW: Revised reference manual. Working paper 2004, Research Department, Bank of Spain. http://www.bde.es.

McElroy, T. (2012). An alternative model-based seasonal adjustment that reduces over-adjustment. *Taiwan Economic Forecast and Policy*, *43*, 33–70.

Wildi, M. (2005). *Lecture notes in economics and mathematical systems*: *Vol. 547*, *Signal extraction: Efficient estimation, unit-root tests and early detection of turning points*. Springer.

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