

Business Cycle Analysis and Zero-Crossings of Time Series: a Generalized Forecast Approach

Marc Wildi

Zurich University of Applied Sciences (ZHAW)

8000 Zurich, Switzerland

marc.wildi@zhaw.ch

June 10, 2025

Abstract

We propose an extension of classic time series approaches to business-cycle measurement called Simple Sign Accuracy (SSA), which addresses zero-crossings of a zero-mean stationary time series. Zero-crossings or sign-changes of the growth-rate of an economic indicator mark transitions between expansion and contraction episodes which can be related to business-cycles. The length or, more specifically, the mean duration between consecutive zero-crossings of a predictor, can be controlled in our approach by subjecting a variation of the classic optimization criterion to a novel ‘holding time’ constraint. The proposed criterion embodies a prediction trilemma which recognizes the fundamental trade-offs between accuracy, timeliness and smoothness (ATS). As a result, the SSA-criterion can address a multiplicity of design priorities in terms of ATS forecast performances, and the classic mean-square error paradigm is obtained as a special case when assigning weight only to the accuracy component. We also show that SSA can be interpreted as an extension of classic smoothing algorithms and that the approach lends itself for customization of existing benchmark predictors. The latter possibility is exploited in a real-time analysis (nowcasting) of the US business cycle, whereby SSA is plugged onto a well-known benchmark to modify the latter’s characteristics in terms of ATS performances.

Keywords: Forecast trilemma, mean-square error, signal extraction, business cycle analysis, zero crossings.

1 Introduction

From a ‘technical’ point of view, business-cycle analysis (BCA) can be related to (and is concerned with) smoothing and filtering. In particular filtering can be associated with a real-time BCA whereby an estimate of the cycle is obtained for *current* time (nowcasting), by relying on a causal one-sided filter applied to one or several economic indicators. Nowcasting poses a challenging prediction problem, because various disparate and to some extent mutually exclusive aspects need to be addressed: accuracy concerns the estimation of the current level of the cycle, above or below the zero-line; timeliness highlights leads and lags or retardation and advancement of the nowcasts; finally, smoothness emphasizes reliability by controlling and containing the number of false ‘noisy’ zero-crossings. Ideally, these three facets, namely accuracy, timeliness and smoothness (ATS), of the prediction problem can be addressed and optimized simultaneously, by a fixed optimization criterion, such that the resulting predictor or cycle nowcast delivers the closest approximation to the true but unobserved level of the cycle, while tracking its sign-changes immediately, without systematic delay, and eventually avoiding the generation of ‘false’ crossings.

Unfortunately, the ATS-terms constitute a prediction trilemma such that when formalized properly, an improvement of one particular term must lead to a loss and deterioration of at least one of the remaining constituents of the trilemma. Some of the underlying two-dimensional trade-offs are intuitively appealing, so for example the dilemma between timeliness and smoothness: a stronger smoothing by a causal design can be obtained by reaching out further into the past of a time series, thus increasing ‘automatically’ its lag. But other aspects are less easy to grasp intuitively: so for example improving simultaneously smoothing and timeliness, to the detriment of accuracy. We provide a new framework to allow users to control and balance all three aspects of the trilemma by expanding existing popular approaches to take account of zero-crossings.

The analysis of zero-crossings of a time series has been pioneered by Rice (1944) who derives a link between the autocorrelation function (ACF) of a zero-mean stationary Gaussian process and its expected number of crossings in a fixed interval. Sign changes or, equivalently, zero crossings of the growth rate of an observed phenomenon, mark transitions between expansion and contraction episodes. In an economic context, these alternating phases of growth and contraction can eventually be associated with a business-cycle (BC), at least if movements are of sufficient magnitude and length. Consecutive zero-crossings of the growth-rate must be spaced away by up to several years to qualify as BC. This mean duration, in turn, asks for a smooth profile of the corresponding indicator: it is precisely this link between the smoothness of a time series and the (mean) duration between its consecutive sign changes, referred to as the *holding time*, which is formalized by Rice (1944).

The classic forecast paradigm, relying on the minimization of the mean-square forecast error (MSE), addresses a single mix of priorities, emphasizing accuracy at costs of timeliness and smoothness. Depending on the application, such a design might suffer from excessive retardation, thereby missing early calls, or from excessive noise leakage, thus generating too many false alarms. For this purpose, Wildi (2005) and subsequently McElroy and Wildi (2019) introduced a forecast approach whose principles root in a formal forecast trilemma but their solution does not accommodate for zero-crossings explicitly, which can be viewed as a shortcoming in some applications. Therefore, we here propose a new so-called Simple-Sign-Accuracy (SSA-) optimization criterion, that controls explicitly for the holding-time of a predictor, i.e., the mean duration between consecutive (zero-)crossings, by relying on an intuitively appealing smoothing constraint based on Rice’s seminal work. Various priorities in terms of accuracy, smoothness and timeliness can be triggered by the selection of a pair of hyperparameters and the classic (MSE-) forecast paradigm is obtained as a special case of our general procedure. Furthermore, we show that SSA can be interpreted as an extension of classic Whittaker-Henderson smoothing and we also demonstrate that the proposed approach allows for customization of benchmark predictors. The latter possibility is exploited in an application of SSA to (real-time) BCA, relying on the well-known

Hodrick-Prescott (1997) (HP-)filter for the benchmark: SSA is asked to track HP in some way optimally, while controlling for retardation and noise leakage of the latter concurrent (one-sided) design. Due to optimal tracking, the interpretation, meaning and economic content of the original HP-benchmark can be carried over to the customized ‘hybrid’ SSA predictor, thus contributing to its explainability. Besides BCA and benchmark customization, our empirical framework proposes simple and intuitively appealing illustrations of the prediction trilemma, including ordinary forecasting, smoothing and signal extraction. All examples can be replicated in an open source SSA-package¹ which extends the application field of SSA-customization to Hamilton’s regression filter, Hamilton (2018), and to the Baxter-King (BK) bandpass filter, Baxter and King (1999).

The optimization criterion is derived in Section (2) together with a customization of BCA tools; Section (3) discusses smoothing, forecasting and a prediction trilemma; Section (4) illustrates an application of the new predictor to BCA and the (real-time) monitoring of the cycle; finally, Section (5) concludes by summarizing our main findings.

2 Simple sign accuracy (SSA-) Criterion

Let $\epsilon_t, t \in \mathbb{Z}$, be Gaussian standard white noise² and let $\gamma_k \in \mathbb{R}$ for $k \in \mathbb{Z}$ be a square summable sequence $\sum_{k=-\infty}^{\infty} \gamma_k^2 < \infty$. Then $z_t = \sum_{k=-\infty}^{\infty} \gamma_k \epsilon_{t-k}$ is a stationary Gaussian zero-mean process with variance $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We consider estimation of $z_{t+\delta}$, $\delta \in \mathbb{Z}$, referred to as the *target*, based on the predictor $y_t := \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$, where b_k are the coefficients of a one-sided causal filter of length L . This problem is commonly referred to as fore-, now- or backcast, depending on $\delta > 0$, $\delta = 0$ or $\delta < 0$, respectively. An extension of this framework to $\tilde{z}_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k}$, where $x_t = \sum_{j \geq 0} \xi_j \epsilon_{t-j}$ is a stationary or non-stationary integrated process³, is proposed in Section (2.3). However, for clarity of exposition and ease of notation we henceforth assume $x_t = \epsilon_t$, acknowledging that straightforward modifications would apply in the case of autocorrelated x_t .

2.1 sign accuracy, MSE and Holding Time

We look for an estimate y_t of $z_{t+\delta}$ such that the probability $P(\text{sign}(z_{t+\delta}) = \text{sign}(y_t))$ is maximized as a function of (b_0, \dots, b_{L-1}) . We now refer to this criterion in terms of *sign accuracy*.

Proposition 1. *Let $\rho(y, z, \delta)$ designate the correlation between y_t and $z_{t+\delta}$. Then, under the above assumptions about ϵ_t, z_t the sign accuracy criterion can be stated as*

$$\max_{b_0, \dots, b_{L-1}} \rho(y, z, \delta) \quad (1)$$

where $\rho(y, z, \delta) = \frac{\sum_{k=0}^{L-1} \gamma_{k+\delta} b_k}{\sqrt{\sum_{k=-\infty}^{\infty} \gamma_k^2} \sqrt{\sum_{k=0}^{L-1} b_k^2}}$ depends on y_t and $z_{t+\delta}$ in terms of their respective filter coefficients b_0, \dots, b_{L-1} and γ_k , $|k| < \infty$.

Proof:

The expression for $\rho(y, z, \delta)$ follows directly from the white noise assumption. Then,

$$\begin{aligned} P(\text{sign}(z_{t+\delta}) = \text{sign}(y_t)) &= 2E[I_{\{z_{t+\delta} \geq 0\}} I_{\{y_t \geq 0\}}] \\ &= 0.5 + \frac{\arcsin(\rho(y, z, \delta))}{\pi} \end{aligned}$$

¹An R-package together with instructions, practical use-cases and theoretical results are to be found at <https://github.com/wiaidp/R-package-SSA-Predictor.git>.

²Since zero-crossings of zero-mean stationary processes are insensitive to the scaling, our approach is insensitive to $\sigma^2 = \text{Var}(\epsilon_t)$: for simplicity, we will assume $\sigma^2 = 1$ if not stated otherwise.

³For a stationary process x_t , ξ_j are the weights of the non-deterministic component in the Wold-decomposition of x_t , such that $\sum_{j \geq 0} \xi_j^2 < \infty$. For an ARMA-process, the weights correspond to its MA(∞)-inversion.

where the last equality is derived from Gaussianity. Neglecting constant terms and considering strict monotonicity of the arcsin function in the interval $] -1, 1[$, the expression in Equation (1) is obtained. \square

Note that signs, zero-crossings or correlations are insensitive to the scalings of y_t or z_t . Also, let $\boldsymbol{\gamma}_\delta := (\gamma_\delta, \gamma_{1+\delta}, \dots, \gamma_{L-1+\delta})'$ denote the classic MSE predictor of the target. Then $\mathbf{b} = \boldsymbol{\gamma}_\delta$ maximizes $\rho(y, z, \delta)$ so that sign accuracy and the classic MSE approach are equivalent down to an arbitrary scaling of the predictor y_t under the above assumptions. Let us now take control over the ‘smoothness’ of y_t and consider the so-called *holding time* defined by

$$ht(y|\mathbf{b}, i) = E[t_i - t_{i-1}]$$

where $t_i, i \geq 1$ are consecutive zero-crossings of y_t , i.e., $t_{i-1} < t_i, t_1 \geq L$, $\text{sign}(y_{t_{i-1}} y_{t_i}) < 0$ for all i and $\text{sign}(y_{t_{i-1}} y_{t_i}) > 0$ if $t_{i-1} < t < t_i$. Under the above assumptions, $ht(y|\mathbf{b}, i) = ht(y|\mathbf{b})$ does not depend on i and is the expected duration between arbitrary consecutive zero-crossings of y_t , see Kedem (1986). Smoothness of y_t is then obtained by constraining \mathbf{b} such that

$$ht(y|\mathbf{b}) = ht_1, \quad (2)$$

where ht_1 is a proper hyperparameter of our design.

Proposition 2. *Let y_t be the above zero-mean stationary Gaussian process. Then the expected holding time $ht(y|\mathbf{b})$ between consecutive zero-crossings of y_t is*

$$ht(y|\mathbf{b}) = \frac{\pi}{\arccos(\rho(y, y, 1))}, \quad (3)$$

where $\rho(y, y, 1) = \frac{\sum_{k=1}^{L-1} b_{k-1} b_k}{\sum_{k=0}^{L-1} b_k^2}$ designates the lag-one autocorrelation of y_t , depending on the latter through the corresponding filter coefficients b_0, \dots, b_{L-1} .

We refer to Kedem (1986) for proof, noting that the expression for $\rho(y, y, 1)$ is derived from the white noise assumption. Equation (3) is remarkable insofar that the ACF at higher lags $k = 2, 3, \dots$ is irrelevant for the holding time. For illustration, we consider the zero-mean AR(1)-process $y_t = a_1 y_{t-1} + \epsilon_t$ for which $\rho(y, y, 1) = a_1$: increasing $\rho(y, y, 1)$ leads to a smoother path of y_t , with longer swings above or below the zero-line and therefore lesser crossings, as implied by the proposition. However, if we tried to address crossings by the lag-two ACF $\rho(y, y, 2) = a_1^2$ instead, this attempt would lead to ambiguity, since a larger $\rho(y, y, 2)$ could lead either to smaller or to larger holding times, depending on the sign of a_1 . In any case, the bijective link between the holding time and the lag-one autocorrelation in Equation (3) implies that the constraint in Equation (2) can be expressed alternatively as

$$\rho(y, y, 1) = \rho_1, \quad (4)$$

where it is assumed that ht_1 and ρ_1 are linked through Equation (3). In the following, we will refer to the ‘holding time’ either in terms of $ht(y|\mathbf{b})$ or $\rho(y, y, 1)$, clarifying our intent in case of ambiguity. The following proposition derives bounds for admissible ρ_1 in Equation (4) and, by extension, for admissible ht_1 :

Proposition 3. *Maximal and minimal lag-one autocorrelations $\rho_{\max}(L), \rho_{\min}(L)$ of y_t as defined above are $\rho_{\max}(L) = -\rho_{\min}(L) = \cos(\pi/(L+1))$. The corresponding moving average (MA) coefficients $b_{\max,k} := \sin\left(\frac{(1+k)\pi}{L+1}\right)$, $k = 0, \dots, L-1$, and $b_{\min,k} := (-1)^k b_{\max,k}$ are uniquely determined down to arbitrary scaling and sign.*

Proof is provided by Davies, Pate and Frost (1974). Maximal and minimal holding times $ht_{\max}(L), ht_{\min}(L)$ of a MA(L)-filter can be obtained by inserting $\rho_{\max}(L), \rho_{\min}(L)$ into Equation (3), see Table (1) for a selection of filter lengths L .

	Length 2	Length 5	Length 12	Length 24	Length 120	Length 240
Max lag one ACF	0.5000	0.8660	0.9709	0.9921	0.9997	0.9999
Min lag one ACF	-0.5000	-0.8660	-0.9709	-0.9921	-0.9997	-0.9999
Max ht	3.0000	6.0000	13.0000	25.0000	121.0000	241.0000
Min ht	1.5000	1.2000	1.0833	1.0417	1.0083	1.0042

Table 1: Maximal and minimal lag-one ACFs and holding times for a selection of filter lengths.

Consider now the sign accuracy criterion in Equation (1) endowed with the holding time constraint in Equation (4):

$$\max_{\mathbf{B}} \rho(y, z, \delta) \quad \left. \begin{array}{l} \\ \rho(y, y, 1) = \rho_1 \end{array} \right\}. \quad (5)$$

In the following, this constrained optimization problem is referred to as *simple sign accuracy* or SSA criterion and we refer to solutions of this criterion by the acronym SSA or SSA(ht_1, δ) to stress the dependence of the predictor on the pair of hyperparameters. The objective function $\rho(y, z, \delta)$ will be referred to as target correlation. Under suitable regularity assumptions, see Theorem (1) in the appendix, an SSA solution exists and is uniquely determined down to an arbitrary scaling term. By refraining from assigning a specific value to this ‘free’ scaling- parameter, we here mark preference and interest for the sign accuracy perspective. In this context, we stay away from establishing links with pure classification methods, such as, e.g., logit-models, because the MSE paradigm remains key for the exposition of forecast tradeoffs⁴. In a BCA context, the hyperparameter ht_1 and, by extension, $\rho_1(ht_1)$, is intuitively appealing and interpretable and its size can be adjusted to match the cycle to be extracted or to improve smoothing upon a benchmark design. Finally, the classic MSE predictor can be replicated by inserting its specific holding time or lag-one ACF into the constraint of the criterion⁵. In summary, the SSA criterion reconciles MSE, sign accuracy and smoothing requirements in a flexible and interpretable way.

2.2 Customization of BCA Tools

For illustration, we apply the SSA criterion to HP and BK filters with two-sided bi-infinite symmetric coefficients or targets $\gamma_{k,HP}$ and $\gamma_{k,BK}$, $|k| < \infty$. We aim at approximating the corresponding filter outputs $z_{t+\delta,HP}$ and $z_{t+\delta,BK}$, where we select $\delta = 0$ (nowcasting), by predictors $y_{t,HP}, y_{t,BK}$ based on one-sided filters $b_{k,HP}$ and $b_{k,BK}$, $k = 0, \dots, 199$, of length $L = 200$, towards the sample end $t = T$. Figure (1) displays truncated finite-length symmetric and classic concurrent filters in the left top (HP) and bottom (BK) panels: we selected $\lambda = 14400$ for HP and $pu = 96, pl = 18$ for the bandpass $[2\pi/pu, 2\pi/pl]$ of BK, matching typical settings for monthly data (Baxter and King (1999) proposed 6-32 quarters). For ease of visual inspection, we shifted the two-sided designs to the right in the figure: in applications, these filters are acausal and centered at $x_{t+\delta}$. The concurrent HP is discussed in Section (4.1) and the concurrent BK is a truncated version of the bi-infinite symmetric filter matching the white noise assumption. Table (2) reports holding times of HP and BK designs, based on Equation (3): a direct comparison suggests that concurrent filters generate markedly more sign changes than the targets they are supposed to nowcast. Excess ‘noisy’ crossings of the former are often clustered in the vicinity of ‘true’ sign changes of the latter, when filter outputs hover over the zero line. Since target crossings are indicative of regime changes of an observed phenomenon, transitioning from expansion to contraction or conversely, we argue that an explicit control of noisy crossings or false alarms, due to an unduly small holding time of the predictors, is a relevant objective in a real-time BCA context.

⁴A logit-model would fit the parameters of a non-linear bounded function of the data to the discrete signs $\text{sign}(z_{t+\delta})$ of the target using of a suitable error measure. On the other hand, the SSA criterion fits a linear filter of the data to the proper observations $z_{t+\delta}$. This latter problem structure is more amenable to numerical optimization and its solution has a smaller variance (increased efficiency) under the posited assumptions.

⁵On request, the optimal (MSE-) scaling of the SSA solution can be obtained as $s := \frac{\sum_{k=0}^{L-1} \gamma_{k+\delta} b_k}{\sum_{k=0}^{L-1} b_k^2}$.

	Concurrent	Target
HP	7.66	59.55
BK	6.76	14.14

Table 2: Holding times of concurrent (left column) and target (right column) HP and BK filters.

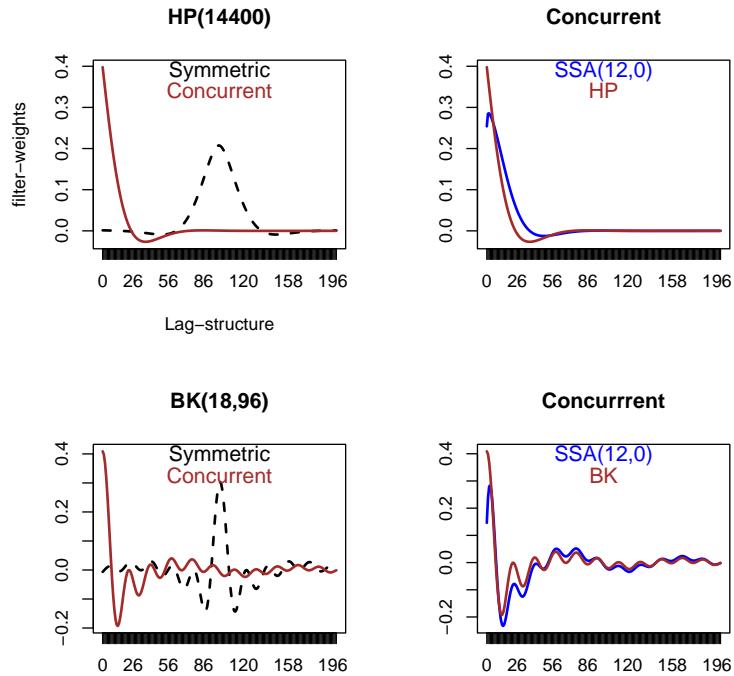


Figure 1: Coefficients of HP, BK and SSA filters. All coefficients are arbitrarily scaled to unit length or unit variance (assuming the data to be standardized white noise).

The SSA(12,0)-designs displayed in Fig.(1) impose a longer holding time $ht = 12$, generating respectively 36.17% and 43.69% less ‘alarms’ than the classic concurrent designs. At the same time, the objective function of criterion (5) ensures that the SSA designs track their respective targets tightly: sign accuracy and correlation are maximized and the MSE is minimized (when relying on the scaling in footnote (5)). The optimal tracking is an important statistical property that is also relevant from an economic point of view because if $z_{t,HP}$ or $z_{t,BK}$ support specific information or economic content, in our case a business cycle indicator, then the corresponding SSA designs inherit this interpretation and meaning from the targets, by optimality of the approximation. To emphasize this point further, we note that the rate of zero-crossings of the concurrent filters could be altered by modifying the hyperparameters λ or pu, pl of the corresponding HP and BK targets; regrettably, this modification would also compromise the cycle specification and therefore the original economic meaning and interpretation of the indicators. On the other hand, as we shall see, SSA minimizes also the rate of zero-crossings in the class of all predictors with the same target correlation (see Corollary (3) in the appendix), and therefore Criterion (5) addresses the problem of noisy false alarms directly and in some way optimally. In summary, the SSA-approach can be ‘plugged’ on a benchmark predictor or nowcaster to modify its characteristics desirably, while preserving the original intent of the analysis. We point out that the selection of the basic framework, to which SSA is applied, is left to the appreciation of the analyst and we

do not advocate in favor of (or against) a specific BCA tool. For its relevance and importance in applications, we here select the HP filter as a target for SSA, see Phillips and Jin (2021) as well as Phillips and Shi (2021) for additional background. Applications of SSA to BK and Hamilton's regression filter, Hamilton (2018), are fleshed out in our SSA-package.

To conclude this section, we briefly relativize the stringency of the posited assumptions. First, we note that y_t or z_t can be nearly Gaussian even if ϵ_t isn't, due to the central limit theorem. Moreover, the constraint could be considered in terms of a generic 'smoothing' requirement, without explicit connection to 'holding times' and criterion (5) remains intuitively appealing irrespective of a strict 'zero-crossing' perspective. Concerning the zero-level of crossings and the underlying zero-mean assumption of the data-generating process, we argue that deviations from the latter hypothesis such as typical for a wide range of economic applications are negligible. Moreover, the smoothing effect entailed by the holding time constraint extends to arbitrary 'non-zero' levels, too. For an illustration of the above claims, Table (3) compares empirical and expected holding times of two SSA designs with $ht_1=3.735$ and $ht_1=10$, which are derived and explained in Section (3.2). Both filters are applied to simulated Gaussian noise, which officiates as the control group, as well as to the last twenty years of daily log-returns of the Standard & Poor's 500 equity index, sampled from 01-01-2000 to 12-31-2021, whereby we consider centered (zero-mean) as well as un-centered data for both series. The table suggests that neither the non-vanishing mean of the log-returns of the equity index nor the presence of volatility clusters, typical of financial data, nor the occurrence of extreme events during the financial crisis or the 'great lockdown', seem to affect the empirical holding times unduly⁶. Additional potentially interesting topics such as heteroscedasticity,

	SSA(3.74,1)	SSA(10,1)
SP-original	3.833 (0.03)	10.248 (0.18)
SP-demean	3.795 (0.03)	10.36 (0.15)
Gauss-original	3.699 (0.03)	9.779 (0.16)
Gauss-demean	3.703 (0.03)	9.71 (0.15)

Table 3: Empirical holding times (sample length divided by the number of zero-crossings) and standard deviations in parentheses of two SSA one-step ahead forecast-filters with theoretical (expected) holding times 3.73 and 10. The filters are applied to simulated Gaussian white noise (control group) as well as to the log-returns of the Standard and Poor's 500 equity index, sampled from 01-01-2000 to 12-31-2021, with and without mean adjustment.

asymmetry or non-linearity of economic time series, in particular during various phases of the cycle, are explicitly discarded from consideration when deriving the SSA predictor (see appendix) and are left as an avenue for future research. We here argue that by plugging SSA on a linear homoscedastic design, such as HP, the approach automatically retains the corresponding classic set of implicit or explicit assumptions and model restrictions. Note, however, that some types of asymmetry, such as different recession and expansion durations, could be traced back in part to a non-vanishing mean of the (differenced) series, which would not affect the proposed analysis. In any case, an extension addressing autocorrelation is proposed in the next section.

⁶The negligible positive biases of the holding times for the equity index can be attributed to extreme observations during high-volatility phases: a single outlier appears as a 'pulse' to the filter which triggers the impulse function at the filter output. Sign-changes of the latter then follow crossings of the former, which are more sparsely distributed.

2.3 Extension to Stationary Processes

In applications, the data to be filtered is typically autocorrelated. Therefore, we first extend the white noise assumption to the case of a stationary x_t in terms of a suitable transformation. Let

$$\begin{aligned} x_t &= \sum_{i=0}^{\infty} \xi_i \epsilon_{t-i} \\ \tilde{z}_t &= \sum_{|k|<\infty} \gamma_k x_{t-k} \end{aligned}$$

be stationary Gaussian processes and designate by ξ_i the weights of the (purely non-deterministic) Wold-decomposition of x_t , such that $\sum_{i=0}^{\infty} \xi_i^2 < \infty$. Then, target and estimate can be formally re-written as

$$\begin{aligned} \tilde{z}_t &= \sum_{|k|<\infty} (\gamma \cdot \xi)_k \epsilon_{t-k} \\ y_t &= \sum_{j \geq 0} (b \cdot \xi)_j \epsilon_{t-j} \end{aligned}$$

where $(\gamma \cdot \xi)_k = \sum_{m \leq k} \xi_{k-m} \gamma_m$ and $(b \cdot \xi)_j = \sum_{n=0}^{\min(L-1,j)} \xi_{j-n} b_n$ are convolutions of the sequences γ_k and b_j with the Wold-decomposition ξ_i of x_t . In the following, it is assumed that $\sum_{k=-\infty}^{\infty} (\gamma \cdot \xi)_k^2 < \infty$ so that \tilde{z}_t is a stationary process: this condition applies, for example, if all sequences are absolutely summable, see Brockwell and Davis (1993). The SSA criterion then becomes

$$\max_{(b \cdot \xi)} \frac{\sum_{k \geq 0} (\gamma \cdot \xi)_{k+\delta} (b \cdot \xi)_k}{\sqrt{\sum_{|k|<\infty} (\gamma \cdot \xi)_k^2} \sqrt{\sum_{j \geq 0} (b \cdot \xi)_j^2}} \quad (6)$$

$$\frac{\sum_{j \geq 1} (b \cdot \xi)_{j-1} (b \cdot \xi)_j}{\sum_{j \geq 0} (b \cdot \xi)_j^2} = \rho_1 \quad (7)$$

which can be solved for $(b \cdot \xi)_j, j = 0, 1, \dots$ (see Theorem (1) in the appendix). The sought-after filter coefficients b_k can then be obtained from $(b \cdot \xi)_j$ by deconvolution or inversion: assuming $\xi_0 \neq 0$ one can start at $j = 0$ with $b_0 = (b \cdot \xi)_0 / \xi_0$ (otherwise the inversion would start at the smallest j such that $\xi_j \neq 0$ and $b_0 = \dots = b_{j-1} = 0$); solving recursively we then obtain $b_{k+1} = \frac{(b \cdot \xi)_{k+1} - \sum_{j=0}^k \xi_{k+1-j} b_j}{\xi_0}$.

We now illustrate criterion (6), assuming that the data x_t conforms to the AR(1) difference equation $x_t = a_1 x_{t-1} + \epsilon_t$, to which the monthly HP filter proposed in Section (2.2) is applied. Figure (2) compares concurrent SSA(12,0) filters for $a_1 = -0.6, 0, 0.6$, thus including the white noise case $a_1 = 0$ in Fig.(1) for reference: top and bottom panels display filter coefficients and amplitude functions, respectively, highlighting early lags and higher frequencies in the right panels. To interpret some of the distinguishing filter patterns in the figure, Table (4) lists the holding times of the original concurrent HP filter when applied to the specified AR(1)-processes (first row). Holding times of the HP filter depend on the data generating process (DGP) and increase markedly with increasing autocorrelation of the data: applying a fixed filter to data with unequal dependence structure can lead to qualitatively different components or cycles. In contrast, SSA(12,0) designs in the second row maintain a fixed holding time $ht = 12$, irrespective of a_1 . For the first two processes in the first two columns, the holding times of HP are smaller than the SSA-specification $ht = 12$ and SSA must increase smoothness over the benchmark. In contrast, the holding time $ht = 14.65$ for the third process exceeds the SSA-specification $ht = 12$ and SSA is asked to generate additional noisy crossings over the benchmark. This atypical demand is reflected by the ripples of the corresponding filter coefficients in Fig. (2). In the frequency domain, the tail behavior of the amplitude function marks control of the rate of zero-crossings: for $a_1 = -0.6$ the filter damps

high-frequency noise most effectively; for $a_1 = 0.6$ increased leakage towards frequency π permits the generation of excess noisy crossings while maintaining optimal tracking of the target by the filter.

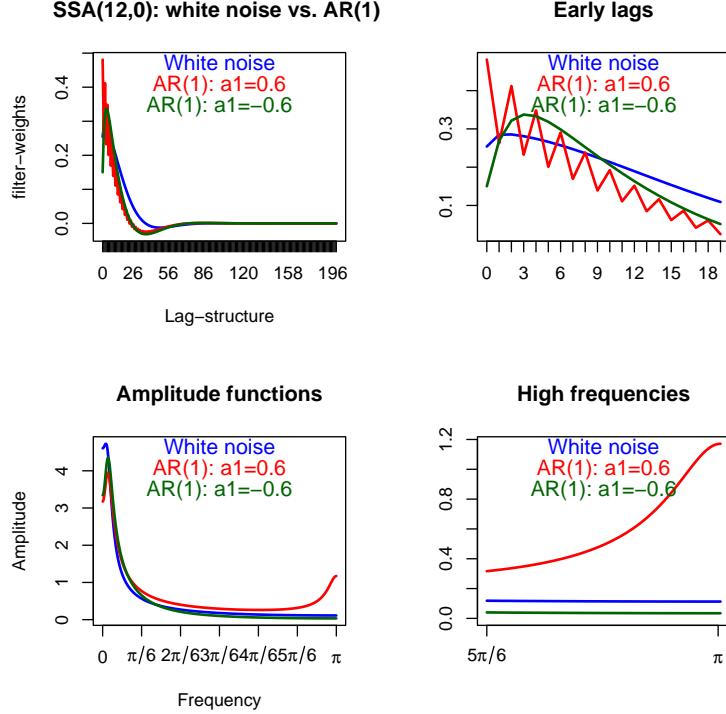


Figure 2: SSA(12,0) based on HP(14400)-target and applied to three different processes. Top left: filters applied to white noise (blue) and AR(1) (red and green); top-right: early lags; bottom-left: amplitude functions; bottom-right: amplitude towards higher frequencies. All filters are arbitrarily scaled to unit length or unit variance, assuming standardized white noise data.

	AR(1)=-0.6	AR(1)=0	AR(1)=0.6
HP	4.02	7.66	14.65
SSA	12.00	12.00	12.00

Table 4: holding times of original HP concurrent filter (first row) and SSA designs (second row) as applied to three different AR(1) processes.

We here omit discussion of the derivation of ξ_k or ϵ_t based on a finite sample x_1, \dots, x_T , see Brockwell and Davis (1993) for background. An analysis of the finite sample estimation error on the holding time of the predictor is provided in the appendix together with an extension to non-stationary integrated processes.

3 Smoothing, Forecasting and a Prediction Trilemma

3.1 SSA vs. Whittaker-Henderson Smoothing

The following proposition asserts that the MSE predictor can be used as an equivalent target specification for the SSA criterion.

Proposition 4. Let $\hat{z}_{t+\delta} = \sum_{k=0}^{L-1} \gamma_{k+\delta} \epsilon_{t-k}$ denote the classic MSE predictor of z_t . Then $z_{t+\delta}$ in criterion (5) can be replaced by $\hat{z}_{t+\delta}$.

Proof

The proof follows from

$$\begin{aligned} \text{Arg} \left(\max_{\mathbf{b}} \rho(y, \hat{z}, \delta) | \rho_1 \right) &= \text{Arg} \left(\max_{\mathbf{b}} \frac{\sum_{k=0}^{L-1} b_k \gamma_{k+\delta}}{\sqrt{\sum_{k=0}^{L-1} b_k^2} \sqrt{\sum_{k=0}^{L-1} \gamma_{k+\delta}^2}} \middle| \rho_1 \right) \\ &= \text{Arg} \left(\max_{\mathbf{b}} \frac{\sum_{k=0}^{L-1} b_k \gamma_{k+\delta}}{\sqrt{\sum_{k=0}^{L-1} b_k^2} \sqrt{\sum_{k=-\infty}^{\infty} \gamma_{k+\delta}^2}} \middle| \rho_1 \right) = \text{Arg} \left(\max_{\mathbf{b}} \rho(y, z, \delta) | \rho_1 \right) \end{aligned}$$

where the notation $\cdot | \rho_1$ signifies that the holding time constraint is imposed and $\text{Arg}(\cdot)$ means the solution of the optimization problem. The expressions to the left and the right of the second equality differ by their scalings $1/\sqrt{\sum_{k=0}^{L-1} \gamma_{k+\delta}^2}$ and $1/\sqrt{\sum_{k=-\infty}^{\infty} \gamma_{k+\delta}^2}$ which accommodate for the different lengths or variances of MSE and bi-infinite target filters. Since both terms are constants, they do not affect the optimization outcome. The remaining two equalities are definitions. \square

This reformulation of the original problem in terms of $\hat{z}_{t+\delta}$ or, equivalently, γ_δ does not refer explicitly to a prediction problem anymore since the new target is causal. Instead, SSA addresses a smoothing problem determined by the holding time constraint. In this context, we could consider any causal filter γ as a potential target: asking for a stronger smoothness by SSA would address ‘noisy’ crossings, often clustered at time points when the target filter hovers over the zero-line. An instructive simple SSA smoothing-problem is obtained by selecting $\gamma = 1$, so that $z_t = x_t$ (identity), and $\delta = -(L-1)/2$ (backcast), assuming L to be an odd number. In this case, the solution y_t of

$$\left. \begin{aligned} \max_{\mathbf{b}} \rho(y, x, \delta = -(L-1)/2) \\ \rho(y, y, 1) = \rho_1 \end{aligned} \right\} \quad (8)$$

aims at tracking $x_{t+\delta}$ while being smoother if $\rho_1 > \rho(x, x, 1)$, the lag-one ACF of the data. Selecting $\delta = -(L-1)/2$ ensures symmetry of the backcast: the coefficients of the causal filter are centered at $x_{t-(L-1)/2}$ and the tails are mirrored at the center point. Let then $\tilde{y}_t := y_{t-\delta}$ denote the so-called SSA smoother, the solution of Criterion (8) shifted forward and centered at x_t . The proposed optimization problem can be considered as an original smoothing algorithm, unrelated to the HP filter because the target $\gamma = 1$ does not refer to a specific benchmark anymore. We can then contrast our approach with classic Whittaker-Henderson (WH) smoothing, see Whittaker (1922), who proposes to solve the following optimization problem for $\mathbf{u} := (u_1, \dots, u_T)$

$$\min_{\mathbf{u}} \left(\sum_{t=1}^T (x_t - u_t)^2 + \lambda \sum_{t=d+1}^T (\Delta^d u_t)^2 \right). \quad (9)$$

The HP filter is obtained by selecting $d = 2$, emphasizing squared second-order differences, i.e., the curvature, in the penalty term. In the case of stationary data, increasing λ typically leads to a longer holding time of u_t but Criterion (8) is more apt at controlling this particular characteristic. For illustration, Fig.(3) displays HP and two different SSA smoothers, all based on $L = 401$, and Table (5) compares their performances. For an identical holding time, SSA1 (blue line in the figure) outperforms HP in terms of target correlation. Interestingly, a dual result applies since for an identical target correlation, SSA2 (violet line in the figure) outperforms HP in terms of holding time, see Corollary (3) in the appendix. However, if mean-square second-order differences are the main gauge for assessing noise suppression, then HP outclasses both SSA smoothers. The observed discrepancies in each one of the reported performance measures seem sufficiently important to ask

	HP	SSA1	SSA2
Holding times	59.580	59.580	75.000
Target correlations	0.205	0.228	0.205
RMS second-order differences	0.096	0.475	0.305

Table 5: HP vs. two different SSA smoothers. SSA1 replicates the holding time of HP and SSA2 replicates its target correlation. Root mean-square second-order differences in the last row refer to standardized white noise data.

for an informed decision: minimizing the rate of zero-crossings, by SSA, or minimizing the curvature, by WH. Eventually, a hybridization obtained by plugging SSA on WH, such as discussed in Section(2.2), could bridge the gap between both smoothing concepts.

To conclude, we broaden the scope of the above comparison. An optimal causal or one-sided SSA smoother can be obtained straightforwardly, by specifying $\delta = 0$, instead of $\delta = -(L - 1)/2$, in Criterion (8). Furthermore, the original Criterion (5) addresses more general estimation or prediction problems than the simple identity-smoothing (8), based on $\gamma = 1$. Finally, SSA can accommodate for the DGP in terms of the MA-inversion of x_t . We argue that if a predictor, or a smoother, has to match sign-changes of a target while keeping control of the alarm rate, determined by zero-crossings, as well as of MSE performances and timeliness characteristics, then the hyperparameters ρ_1, δ of SSA, as well as the underlying optimization principle, address facets of that problem in a more nuanced way than λ for the WH smoother and, incidentally, for the HP filter.

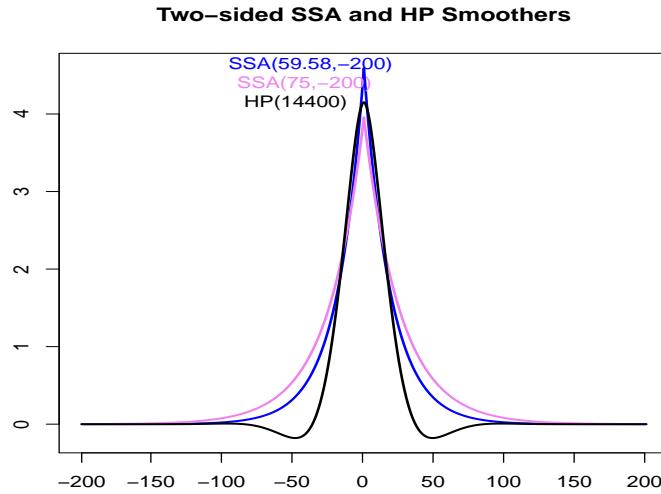


Figure 3: Coefficients of two-sided SSA and HP smoothers of length 401, arbitrarily scaled to unit length: the first SSA design (blue line) replicates the holding time of HP, the second SSA design (violet line) replicates the tracking-ability or target correlation of HP.

3.2 Forecasting and a Prediction Trilemma

For an introduction and an illustration of the proposed concepts, we now briefly consider a simple one-step-ahead forecast exercise of an MA(2)-process

$$z_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$$

where $\gamma_k = 1$, $k = 0, 1, 2$ and $\delta = 1$. Note that the exercise slightly departs from classic time series analysis since the MA-process is non-invertible and because we assume ϵ_t to be known, which is not a limitation, see Section (2.3). For comparison purposes we compute three different SSA forecast filters $y_{ti}, i = 1, 2, 3$ for z_t : the first two are of identical length $L = 20$ with dissimilar holding times $ht = 3.74$ and 10 ; the third filter deviates from the second one by selecting $L = 50$; the holding time of the first filter matches the lag-one autocorrelation of z_t and is obtained by inserting $\rho(z, z, 1) = 2/3$ into Equation (3); the second holding time $ht = 10$ is sufficiently different in size to reveal some of the salient features of the proposed approach. In addition, we also consider the MSE forecast $\hat{z}_{t+1}^{MSE} = \epsilon_t + \epsilon_{t-1}$, as obtained by classic time series analysis, as well as a trivial ‘lag-by-one’ forecast $\hat{z}_{t+1}^{lag\ 1} = z_t$, see Fig. (4) (an arbitrary scaling scheme is applied to SSA filters).

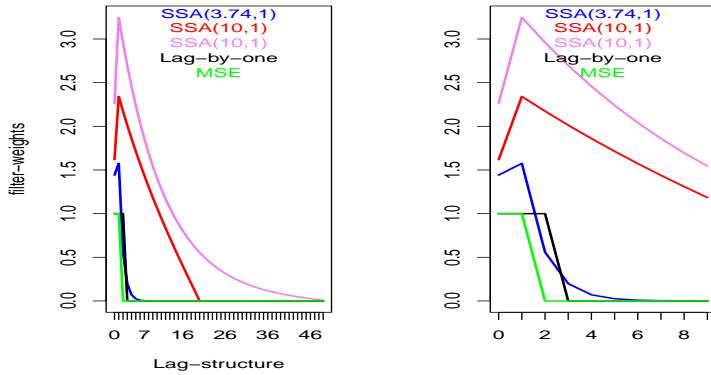


Figure 4: Coefficients of MSE-, SSA- and lag-by-one forecast filters with arbitrarily scaled SSA designs. All lags (left panel) and first ten lags (right panel).

Except for the MSE (green) all other filters rely on past ϵ_{t-k} for $k > q = 2$ which are required for compliance with the holding time constraint (stronger smoothing). For a fixed filter length L , a larger holding time ht asks for a slower zero-decay of filter coefficients (blue vs. red lines) and for fixed holding time ht , a larger L leads to a faster zero-decay but a long tail of the filter (red vs. violet lines). The distinguishing tips of the SSA filters at lag one in this example are indicative of one of the two implicit boundary constraints, namely $b_{-1} = 0$, see Theorem (1) in the appendix. Note that the ‘lag-by-one’ forecast (black) has the same holding time as the first SSA filter (blue) so that the latter should outperform the former in terms of sign accuracy or, equivalently, in terms of target correlation with the shifted target, as confirmed in Table (6). MSE

	SSA(3.74,1)	SSA(10,1)	SSA(10,1)-long	Lag-by-one	MSE
Target correlation	0.786	0.386	0.388	0.667	0.816
Empirical holding times	3.735	10.000	10.000	3.735	3.000
Empirical sign accuracy	0.788	0.626	0.627	0.732	0.804

Table 6: Performances of MSE and lag-by-one benchmarks vs. SSA: All filters are applied to a sample of length 1000000 of standardized Gaussian noise. Empirical holding times are obtained by dividing the sample length by the number of zero-crossings. The two columns referring to SSA(10,1) correspond to filter lengths 20 (first) and 50 (second).

outperforms all other forecasts in terms of correlation and sign accuracy but it loses in terms of smoothness or holding time; SSA(3.74,0) outperforms the lag-by-one benchmark; both SSA(10,0) loose in terms of sign accuracy but win in terms of smoothness and while the profiles of longer and shorter filters differ in Figure (4), their respective performances are virtually indistinguishable

in Table (6), suggesting that the selection of L is not critical (assuming it is at least twice the holding time). The table also illustrates the tradeoff between MSE- or sign accuracy performances of optimal designs, in the top and bottom rows, and smoothing performances in the middle row (an explicit formal link can be obtained but is omitted here). Table (7) allows for a more detailed analysis based on a finer grid of holding time values. In a business cycle application, MSE per-

	ht=4	ht=4.5	ht=5	ht=5.5	ht=6	ht=7	ht=8	ht=9	ht=10
Correlation	0.77	0.72	0.68	0.64	0.60	0.53	0.47	0.43	0.39
Emp. ht	4.00	4.50	5.00	5.50	6.00	7.00	8.00	9.00	10.00
Sign accuracy	0.78	0.76	0.74	0.72	0.70	0.68	0.66	0.64	0.63

Table 7: Tradeoff: effect of the holding time on target correlation (first row) and sign accuracy (last row) for fixed forecast horizon.

formances or, equivalently, the target correlation (first row in the table), are related to assessing the level of the cycle, i.e., its precise value above or below the zero-line, whereas the holding time (second row) emphasizes performances at the zero-line, specifically, with the intent to address the number of random crossings due to noise-leakage. The SSA framework allows to accord the design of the predictor with the particular purpose of the analysis, by a suitable balance of the observed tradeoff. Different filters could be used, in isolation or combination, for measuring the level with higher accuracy, but reduced smoothness, or for assessing sign changes in the growth rate more reliably. In the latter case, given a specified loss in target correlation, the dual reformulation of the SSA criterion (see Corollary (3) in the appendix) states that the resulting filter effectively minimizes the rate of zero-crossings or alarms. Formally, a decision for one or several interesting designs could be based on re-computing the above ‘tradeoff-table’ with the SSA-package for the particular prediction problem at hand, accounting for the DGP (Wold-decomposition) as well as for the target, as we did above.

In a final step, we allow the previously fixed forecast horizon $\delta = 1$ to vary, see Fig.(5) for illustration.

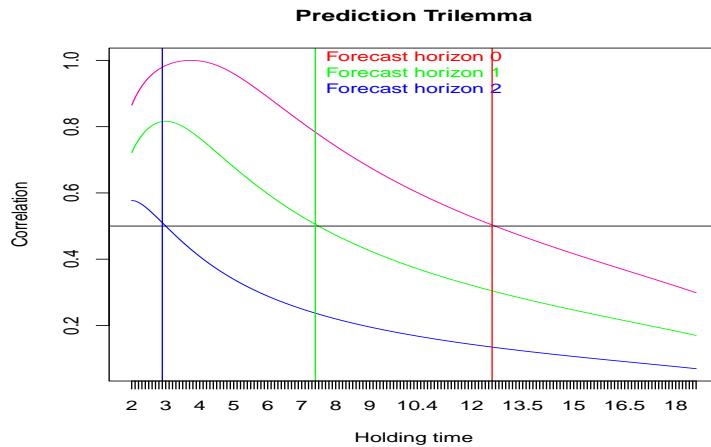


Figure 5: Target correlations of the SSA predictor as a function of the forecast horizon and the holding time.

For each $\delta = 0, 1, 2$, the figure displays the target correlation for given holding times on the abscissa. The peak of the correlation for a given δ appears at a particular holding time value which corresponds to the classic MSE predictor for that forecast horizon. To the right of the

peak, SSA generates fewer crossings (stronger smoothing than MSE) and to its left more; on both sides, SSA maximizes the target correlation, subject to the imposed holding time constraint ht_1 on the abscissa; in its equivalent dual form, SSA maximizes the holding time for given target correlation on the right of the peak; finally, the correlation curve for $\delta = 1$ replicates entries in Table (7). For a fixed target correlation, a larger ht_1 (increased smoothness) is functionally related to a smaller δ (reduced timeliness). Specifically, consider the three pairings (ht_{1i}, δ_i) , $i = 1, 2, 3$, with values $(2.9, 2)$, $(7.4, 1)$ and $(12.6, 0)$ marked by vertical lines in the figure: the corresponding $SSA(ht_{1i}, \delta_i)$ -predictors y_{ti} have a fixed correlation $\rho(y_{ti}, z, \delta_i) = 0.5$ with the target $z_{t+\delta_i}$, marked by the horizontal black line which intersects the curves at the corresponding holding times ht_{1i} , $i = 1, 2, 3$. Fig.(5) generalizes the dilemma in Table (7) to a *prediction trilemma*, by allowing timeliness, embodied by δ , to become a separate structural element, or hyperparameter, of the estimation problem, together with ht_1 . For a particular target $z_{t+\delta_0}$, the pair (ht_1, δ) spans a two-dimensional space of predictors $SSA(ht_1, \delta)$ and classic MSE performances can be replicated by selecting $\delta = \delta_0$ and $ht_1 = ht_{MSE}$, the holding time of the MSE predictor. However, alternative priorities in terms of timeliness or smoothness can be triggered by screening the two-dimensional predictor space and our SSA-package can be used to assess an optimal balance of the constituents of the trilemma for general prediction problems.

To conclude this topic we provide advice on the selection of hyperparameters (ht_1, δ) of the SSA predictor for a target $z_{t+\delta_0}$. To determine ρ_1 or, equivalently, ht_1 , it is advisable to compute reference values, such as $ht_{two-sided}$ and ht_{bench} , the holding times of the acausal target and a benchmark predictor, such as the classic MSE-design or, in our case, the concurrent HP filter. Typically, $ht_1 > ht_{bench}$ to lessen the rate of noisy alarms when compared to the benchmark. Also, $ht_1 < ht_{two-sided}$ avoids smoothing out target crossings. In applications, reducing the alarm rate of the benchmark by up to 50% is often acceptable when gauged against losses of timeliness or target correlations. Timeliness addresses the relative lead or lag of the SSA predictor with respect to the benchmark. The nowcast $\delta = 0$ is a plausible setting for a real-time BCA framework. Typical selections for forward-looking designs in the context of monthly data are $\delta = 6, 12, 18, 24$. We can now turn our attention to business cycle analysis and apply SSA to the HP filter.

4 Application to BCA

We here apply the SSA design to monthly industrial production indices of various countries with long business cycle histories and benchmark performances against the HP filter, see Hodrick and Prescott (1997). We refer to the two-sided symmetric HP filter as the ‘target’ z_t which must be nowcasted at the current sample-end, i.e., $\delta_0 = 0$. For that purpose, we consider five concurrent designs, namely two classic HP filters, HP gap and HP trend, as well as three SSA designs based on distinct hyperparameter settings for ρ_1, δ , see Section (4.1) for reference. Our implementation of SSA in this application emphasizes simplicity and robustness. In particular, we do not fit models to the data, assuming log-returns to be white noise⁷.

4.1 SSA- and Hodrick Prescott Filters

The HP filter is widely used to estimate trends and cycles of economic time series. It can be interpreted as an optimal MSE-signal extraction filter for the trend in the smooth trend model, see Harvey (1989). Conceptually, this results in ‘implied’ models for the cycle and the trend, such

⁷Growth rates of a wide range of economic time series are in accordance with our simplifying assumption, see the so-called ‘typical spectral shape’ of an economic variable in Granger (1966). Indeed, economic theory suggests that stock prices, futures prices, long-term interest rates, oil prices, consumption spending, inflation, tax rates, or money supply growth rates should follow (near) martingales, see Fama (1965), Samuelson (1965), Sargent (1976), Hamilton (2009), Hall (1978) and Mankiw (1987). Finally, log-returns of the monthly US industrial production index (INDPRO) are compatible with simple AR(1)- or MA(1)-model specifications, with parameter estimates close to zero. A detailed analysis can be found in our SSA-package together with an extension to stationary processes based on Section (2.3).

that applying the HP filter results in MSE optimal estimates. In this framework, the bi-infinite symmetric expansion of the filter is obtained as

$$(\gamma_{|k|} B^k)_{|k|<\infty} = \frac{1}{1 + \lambda(1 - B)^2(1 - B^{-1})^2} \quad (10)$$

where λ is a ‘smoothing’ hyperparameter and where B, B^{-1} are backward and forward operators. The implicit data-generating process follows an ARIMA(0,2,2)-specification whose MA-coefficients are determined by λ , see McElroy (2006). In finite samples, the filter behaves differently toward the boundaries of the data, where the symmetry is lost and an exact finite sample representation of the concurrent HP trend filter, denoted as $b_k^{HP-trend}$, can be found in McElroy (2006) and Cornea-Madeira (2017). Business-cycle analysis concerns the deviation of an indicator from its trend which is sometimes referred to as the output-gap. Accordingly, the classic HP gap filter is defined by

$$b_k^{gap} := \begin{cases} 1 - b_0^{HP-trend} & k = 0 \\ -b_k^{HP-trend} & k > 0 \end{cases} \quad (11)$$

The output-gap embodies the concept of trend-reversion, assuming that the natural equilibrium of the economy lies along an abstract growth path, called potential output, to which the economy reverts after shocks. In Equation (11), the gap relies on HP-trend as a surrogate for the latent trend. Instead, we propose a *growth-cycle* design that attempts to match (in real time) the National Bureau of Economic Research (NBER) business-cycle chronology. This chronology serves as a yardstick against which our designs can be benchmarked. Specifically, the cycle is obtained by applying HP-trend to *first differences* of the original economic indicator. Furthermore, SSA can ameliorate timing and noise suppression of the benchmark HP filter.

In summary, HP-gap, as applied to *levels*, and HP-trend, as applied to *differences*, formalize two different and complementary concepts of the business-cycle: the former is an economic measure of the difference between the actual output of an economy and its potential output; the latter is a measure of the magnitude (and the sign) of its growth-trend. To proceed to meaningful comparisons, we then need to transform the respective filters such that both can be applied to the same time series, namely either data in levels or data in first differences. For HP-gap we propose a modified version $b_k^{\Delta gap}$ with the property

$$\sum_{k=0}^{L-1} b_k^{\Delta gap} \Delta x_{t-k} = \sum_{k=0}^{L-1} b_k^{gap} x_{t-k}$$

where Δx_t are first differences of a time series x_t , see Section (2.3). Note that filter outputs of original and modified gap-filters are strictly identical but their input series differ. One can verify that $b_k^{\Delta gap} = \sum_{i=1}^k b_i^{gap}$ and that the coefficients decay towards zero for increasing lag, see Section (2.3) and McElroy and Wildi (2020)⁸. A similar transformation $b_k^{\Sigma,trend}$ can be applied to HP-trend with the property

$$\sum_{k=0}^{L-1} b_k^{trend} \Delta x_{t-k} = \sum_{k=0}^L b_k^{\Sigma,trend} x_{t-k}$$

$$\text{where } b_k^{\Sigma,trend} = \begin{cases} b_0^{trend} & k = 0 \\ b_k^{trend} - b_{k-1}^{trend} & 0 < k \leq L-1 \\ -b_{L-1}^{trend} & k = L \end{cases} .$$

⁸ b_k^{gap} is a highpass design with the property that $\sum b_k^{gap} = 0$ which follows from the definition (11). Therefore $b_k^{\Delta gap} = \sum_{i=1}^k b_i^{gap} \rightarrow 0$ for increasing k .

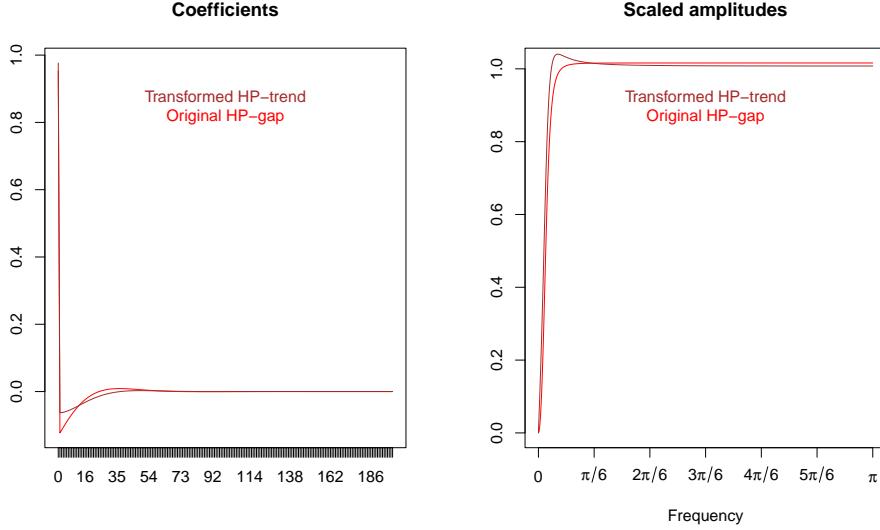


Figure 6: Original concurrent HP-gap (red) and modified HP-trend (brown): scaled coefficients (left panel) and amplitude functions (right panel). Both filters are applied to data in levels.

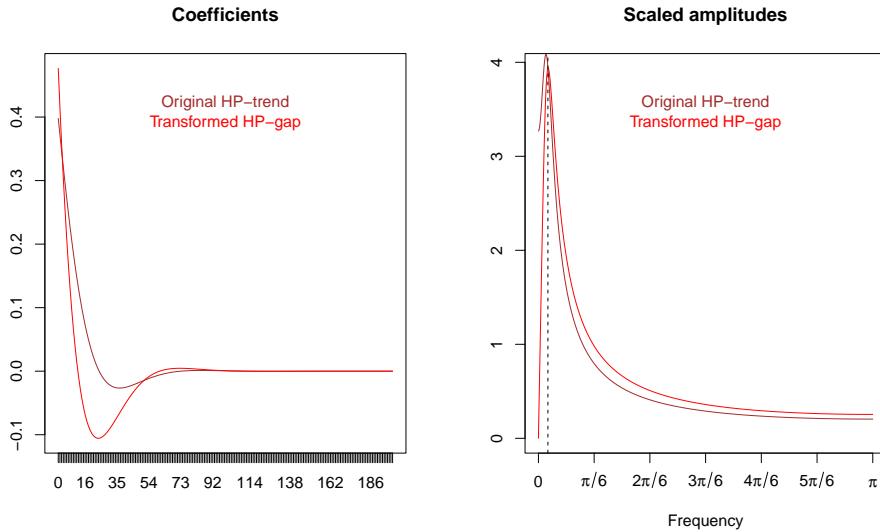


Figure 7: Original HP-trend (brown) and modified HP-gap (red): scaled coefficients (left panel) and amplitude functions (right panel). The dashed vertical line marks the peak of the gap's amplitude. Both filters are applied to data in first differences.

Following Morten and Uhlig (2002), we select $\lambda = 14400$ (monthly data) for the HP-filter. Fig.(6) compares (scaled) original HP-gap and modified HP-trend: both filters are applied to data in levels; Fig.(7) displays (scaled) modified HP-gap and original HP-trend: both filters are applied to data in first differences. The filters in the first graph are highpass designs and their similar profiles might suggest similarity of their respective filter outputs. However, the amplitude functions in the second graph reveal that, in terms of first differences, the HP-trend is a *lowpass* and

(transformed) HP-gap is a *bandpass*, with a vanishing amplitude at frequency zero. To highlight this point, we apply both filters to an artificial time series: the differenced series consists of a time dependent deterministic level component overlaid with noise, see Fig.(8) top panels. The bottom panels illustrate that the filter outputs based on differenced data are the same as when based on data in levels, after suitable transformation(s).

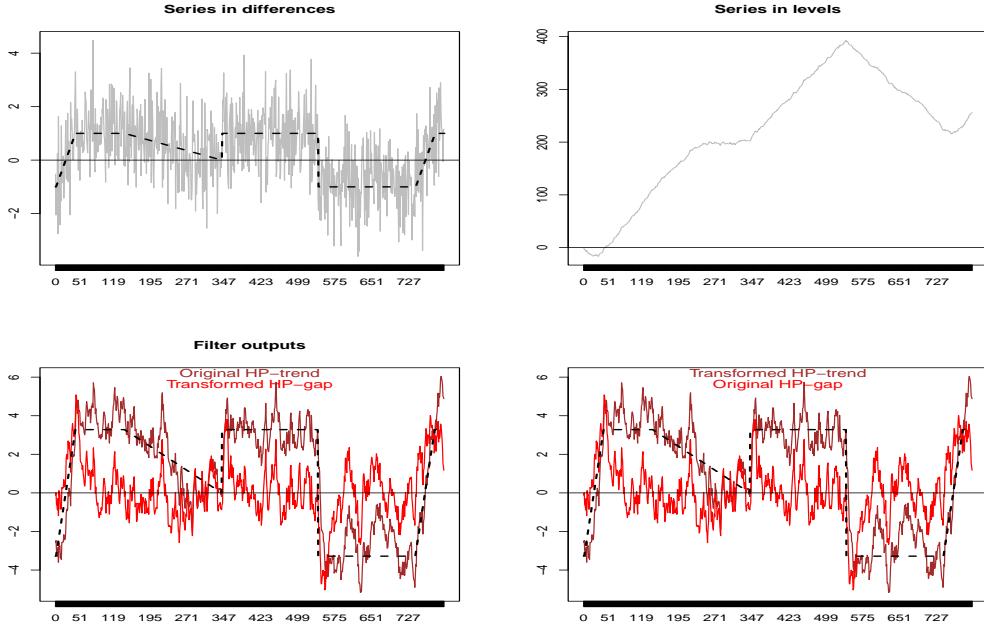


Figure 8: Artificial data in differences (left top) and levels (right top): variable growth rate (dashed line) overlaid with Gaussian noise. Filter outputs of original HP-trend (brown) and modified HP-gap (red), as applied to differences (bottom left); transformed HP-trend (brown) and original HP-gap (red) applied to levels (bottom right). The corresponding filter outputs in left and right (bottom) panels are strictly identical.

The HP-trend can track growth-rate shifts which are a salient feature of the data. In contrast, the HP-gap dismisses the ‘signal’ and fluctuates around the zero line: irrespective of the actual growth-rate, the gap is evenly distributed around zero. We argue that the output-gap in this simulation is mainly an expression of the bandpass characteristic of the filter in the sense that the mean-duration of the ‘cycle’ can be related to the periodicity of the peak-amplitude frequency of the filter (dashed vertical line in Fig.(7)), whose location is reliant upon λ . The peak-amplitude frequency corresponds to a duration of 70.47 (months) with half-cycle lengths of three years so that we expect to observe multiple cycles during longer expansion episodes of the (US) economy, see Section (4.2). The growth-cycle, on the other hand, is a manifestation of the changing level of the data as tracked by HP-trend (brown line) and the additional ‘cyclical’ movements of the filter about the signal are unwanted (noisy) components attributable to high-frequency leakage of HP-trend: these components, which correlate with the gap, are of no particular interest for the determination of the growth-cycle. The observed output-gap in this example can be assimilated to a filter artifact (spurious cycle) which cannot be traced back to a salient feature of the data; in contrast, the growth-cycle extracts a pertinent and meaningful data-feature. From a growth-cycle perspective, HP-gap cancels the relevant signal and the estimated gap is (undesirable) ‘noise’; conversely, in a gap-perspective the low-frequency part of the growth-cycle is an undesirable nuisance that must be canceled. If changes of the growth-rate of the original time series are sufficiently important, due to the alternation of economic contraction and expansion episodes, then the result-

ing magnitude of the systematic level-changes of HP-trend eventually dominates the noisy ‘cycle’ seeping through the filter, due to leakage, and zero-crossings of the filter output can be associated with transitions between contractions and expansions of the series in levels. For illustration, the left bottom panel of Fig.(8) suggests that HP-trend is free of zero-crossings once the level-shift is sufficiently important. In contrast, the unfiltered time series in the left top panel is still subject to multiple crossings at any level of the true growth-rate.

The unwanted ‘cyclical’ noise component of the HP-trend generates additional false zero-crossings at the transition between positive and negative growth episodes and SSA can address noise leakage as well as retardation of the one-sided filter in some way optimally. For the BCA application we then select $ht_1 = 12$, which matches roughly the mean-duration of recessions (see also the closing discussion in Section (4.2)) and $ht_1 = 7.66$ which is the expected holding time of the benchmark HP(trend)-filter, as based on Equation (3). Timeliness is addressed by selecting either $\delta = 0$ (nowcast) or $\delta = 18$ (forecast), see Fig.(11) for further analysis and keep in mind that the proper target is a nowcast, i.e., $\delta_0 = 0$ is fixed. Our selection of the hyperparameters underweights MSE performances in favor of noise suppression and lead or advancement in the prediction trilemma of Section (3.2).

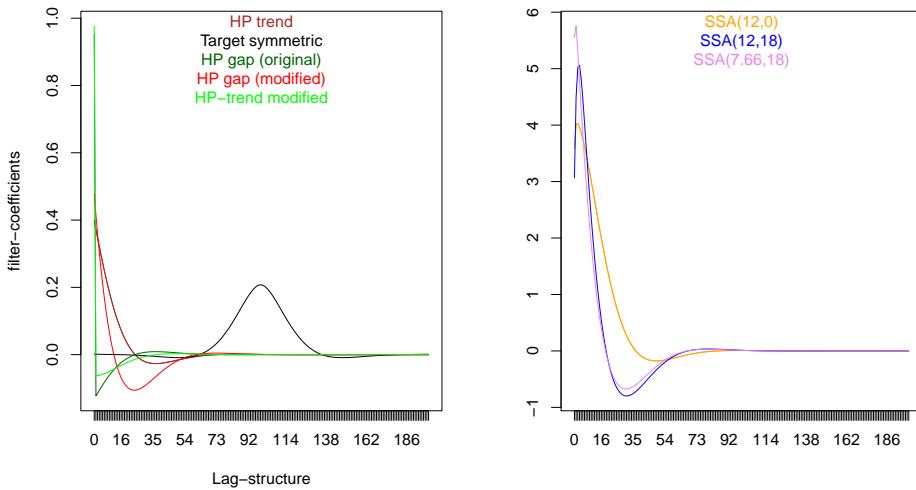


Figure 9: HP concurrent (left panel) and SSA concurrent filters (right panel). All filters are arbitrarily scaled to unit length.

The coefficients of the specified concurrent and target filters are displayed in Fig.(9): for simplicity of exposition all filters have length $L = 200$ and are normalized to unit variance ($\frac{1}{L} \sum_{k=0}^{L-1} b_k^2 = 1$). The characteristic tips of the SSA filters are indicative of the boundary constraint $b_{-1} = 0$, see Theorem (1) in the appendix. The first two (orange and blue) have identical holding times (smoothness) but we expect different lead-lag properties (timeliness); the third (purple) has a shorter holding time matching HP trend. Fig.(10) compares amplitudes and phase-lags⁹ of all relevant concurrent filters.

⁹The phase-lag at a given frequency ω measures the shift, in time-units, between output and input of the filter when fed with a sinusoidal of that frequency.

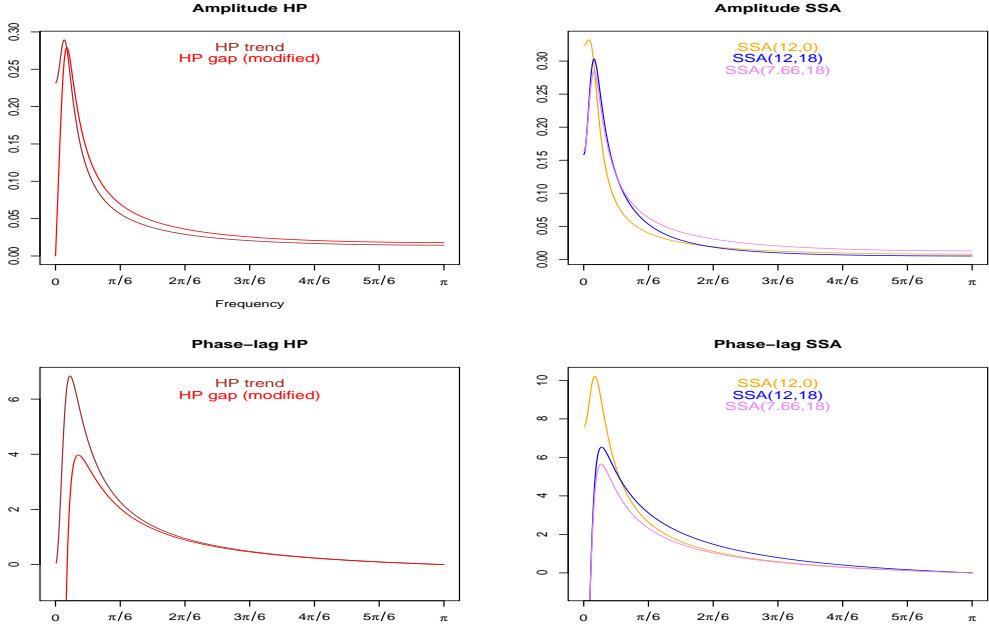


Figure 10: Amplitude and phase-lag functions of modified HP gap (red), HP trend (brown), SSA(12,0) (orange), SSA(12,18) (blue) and SSA(7.66,18) violet)

All filters, except modified HP gap, are lowpass designs. Amplitude functions of SSA with larger holding times (orange and blue) are smallest at higher frequencies, due to stronger smoothing. The phase-lag of HP gap is the smallest overall; SSA(7.66,18) (orange) outperforms HP trend uniformly at business cycle frequencies; SSA(12,18) outperforms HP trend only at lower cycle-frequencies, see Fig.(11).

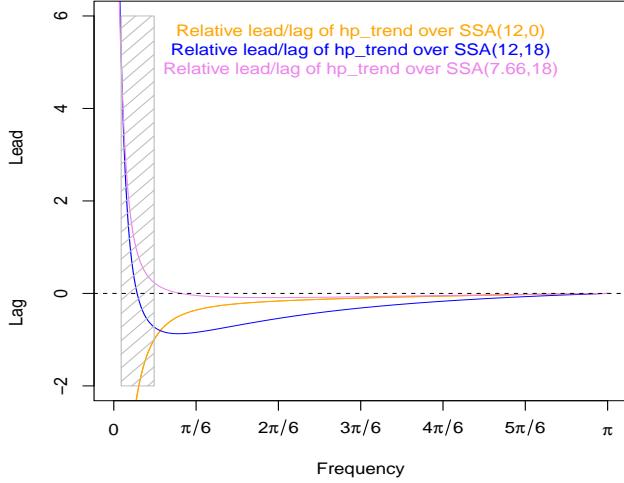


Figure 11: Difference of phase-lags of SSA vs. HP trend. Positive values signify a lead of SSA at the corresponding frequency. Business cycle frequencies, i.e., periodicities between two and ten years are highlighted by the shaded area.

An application of SSA and HP trend to simulated Gaussian noise leads to mean-shift or τ -statistics summarized in Table (8) (the mean-shift or τ -statistic is discussed in the appendix: it measures the shift of two competing filter outputs at zero-crossings and a positive mean-shift implies a corresponding lead or left-shift of the reference-filter, averaged over all crossings). We

	SSA(12,0)	SSA(12,18)	SSA(7.66,18)
Mean-shift	-1	0	1

Table 8: Mean-shift (τ -statistic) of HP trend as referenced against SSA based on an application to Gaussian white noise: positive numbers suggest a lead or left-shift by SSA

now apply the filters to economic data: a detailed analysis of the US business cycle is proposed in Section (4.2) and summary-statistics for a selection of additional countries are provided in Section (4.3).

4.2 Application to the US-Industrial Production Index

We consider an application of the proposed concurrent filters to first differences ΔI_t of the (log-transformed) monthly US industrial production index I_t plotted in Fig.(12): the series effectively starts in 1919-02-01 (FRED database) but we display shorter samples for ease of visual inspection.



Figure 12: Monthly US industrial production index (INDPRO) and recession episodes as dated by the National Bureau of Economic Research, NBER (shaded)

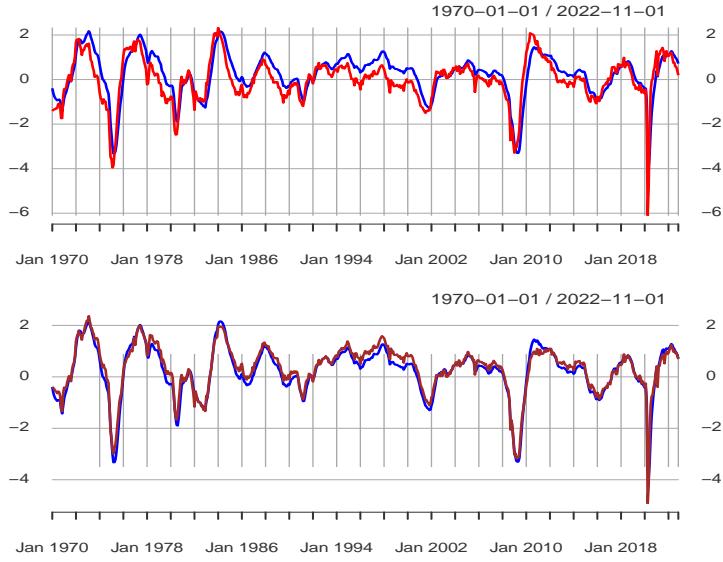


Figure 13: Filter outputs: HP gap (red) and SSA(12,18) (blue) in the top panel; HP trend (brown) and SSA(12,18) (blue) in the bottom panel.

Filter outputs, scaled to unit variance, are displayed in Fig.(13). The figure illustrates that the bandpass HP gap (top panel) generates excessively many crossings, see Table (9), as well as false systematic sign changes, i.e., spurious cycles, during longer up-swings covering the great moderation, from the early nineties up to the financial crisis, see also Fig.(14), left panels. In contrast, the lowpass designs track the longer cycle dynamics better. All filters indicate a slowdown of industrial production in 2015 and 2016, at a time when the price for crude oil declined sharply, hence affecting petrol extraction as well as collateral industrial activity in the US. A potential

advantage of imposing stronger smoothness can be seen in Fig.(14), right panels, which highlight the pandemic ‘great-lockdown’ crisis: zero-crossings of the SSA design are fewer and dynamics are less noisy than the benchmark, which facilitates a real-time assessment of economic conditions, at least in the industrial sector. The bottom-right panel illustrates the potential right-shift or lag of SSA(12,0) at zero-crossings, as measured by τ in the appendix. Table (9) compares timeliness and smoothness performances in terms of τ and number of sign changes: HP gap and SSA(7.66,18) outperform in terms of timeliness followed by SSA(12,18), HP trend and SSA(12,0); but the latter SSA(12,0) outperforms in terms of smoothness, followed by SSA(12,18), SSA(7.66,18), HP trend and finally HP gap. These rankings mostly conform with the diagnostics established in the previous section, based on amplitude and phase-lag functions, and they express the tradeoff entailed by the prediction trilemma in Section (3.2). Based on theoretical as well as empirical evidence, we now discard HP gap, subject to spurious cycles, as well as SSA(12,0), subject to a relative lag.

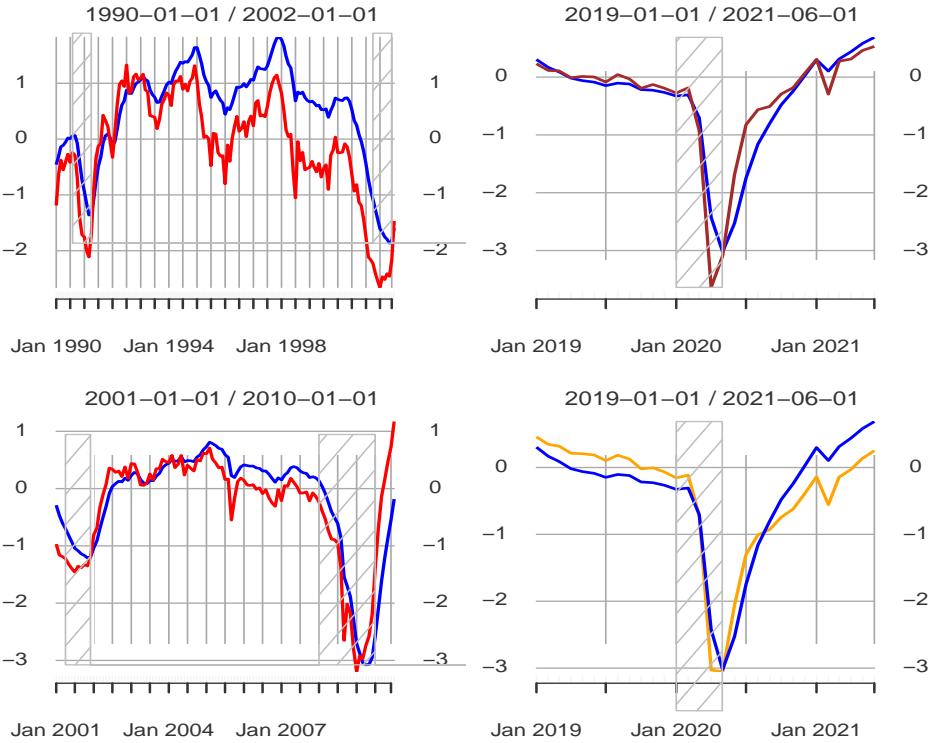


Figure 14: Filter outputs: Modified HP gap (red), HP trend (brown), SSA(12,18) (blue) and SSA(12,0) (orange). From 1990-2002 (top left panel, great moderation), 2000-2010 (bottom left panel) and 2019-2021 (right panel: great lockdown). Shaded areas mark recessions as dated by the NBER.

	SSA(12,18)	HP trend	Modified gap	SSA(12,0)	SSA(7.66,18)
Mean-shift	0	1	-1	1	-1
Number of crossings	44	60	90	40	58

Table 9: Mean-shift (τ -statistic), as referenced against SSA(12,18), and number of crossings for US-INDPRO from 1935-09-01 to 2022-11-01: positive mean-shifts signify a left-shift or lead by SSA(12,18)

To conclude our analysis of the US business cycle, we note that the empirical holding time

$\hat{ht} = 24$ of SSA(12,18) exceeds the expected number $ht_1 = 12$. This discrepancy is mainly due to positive autocorrelation (the ‘cyclical’ growth rates of the series are smoother than white noise) and the observed effect could be addressed by the extension in Section (2.3), fixing the link between expected and empirical holding times. Also, the mean-length of full expansion-recession cycles declared by the NBER amounts to 65 months, i.e., 5 years over a time span from 1935-09-01¹⁰ to 2022-11-01. In comparison, the empirical holding time of the SSA filter suggests a mean cycle length of 48 months, or four years. Therefore, additional fine-tuning of ht_1 might be necessary to match cycle-lengths. Note, however, that some of the downturns of the industrial production index, such as in 2015 and 2016, do not classify as economic recessions and therefore a direct comparison of corresponding ‘cycles’ is subject to caution.

4.3 Multi-National Perspective

We here extend the above framework ‘as is’, without re-tuning HP or SSA designs, to a selection of countries, restricting attention to mature and large economies with correspondingly long and stable cycle histories, ideally differing from the US. Specifically, we here consider Brazil, France, Germany, Italy, Japan, Korea, Spain, United Kingdom and Euro Area 19 countries (we relied on external economic advice for selecting the series according to the cited criteria). All (non-US) series were obtained from the Main Economic Indicator (MEI) database of the OECD, whereby the common sample period ranges from Jan-1980 to Nov-2022¹¹. Table (10) lists country-specific performances but we also provide aggregate performances, obtained by concatenating all series, see Table (11). Our results confirm expectations, as entailed by the selected hyperparameters, and the two selected SSA designs are representative of two particular non-exhaustive research priorities. While the lag of HP trend against the reference SSA(12,18) is statistically insignificant, the latter

	SSA(12,18)	HP trend	SSA(7.66,18)
Brasil shift	0	0	-1
Brasil number crossings	32	46	46
Spain shift	0	0	-1
Spain number crossings	35	71	65
US shift	0	1	-1
US number crossings	44	60	58
Japan shift	0	0	0
Japan number crossings	32	56	54
France shift	0	0	0
France number crossings	38	63	67
Germany shift	0	0	0
Germany number crossings	26	48	52
Italy shift	0	0	0
Italy number crossings	27	48	46
Korea shift	0	-1	0
Korea number crossings	23	37	35
United.Kingdom shift	0	1	0
United.Kingdom number crossings	30	33	41
Euro.area..19.countries. shift	0	0	-1
Euro.area..19.countries. number crossings	17	32	30

Table 10: Number of crossings and mean-shift, as referenced against SSA(12,18): a positive number means a corresponding lead.

outperforms the former in terms of smoothness. The lead of SSA(7.66,18) over the reference

¹⁰The earliest observations are skipped due to filter initialization.

¹¹The series were retrieved under the ‘Total industry excluding construction’ banner. Specifically: Production > Industry > Total industry > Total industry excluding construction.

	SSA(12,18)	HP trend	SSA(7.66,18)
Mean-shift over countries	0.00	0.14	-0.44
t-test for time-shift	0.00	1.25	-9.61
Total number of crossings	304.00	494.00	494.00

Table 11: Aggregate mean timeliness and smoothness performances obtained by concatenation into a single long series. Shifts at zero-crossings i.e. τ -statistics are referenced against SSA(12,18): positive shifts indicate a lead or left-shift of the reference; t-statistics for the significance of the lead or lag are reported in the middle row

SSA(12,18), and thus over the benchmark, is strongly significant: for similar smoothness, SSA outperforms HP trend in terms of timeliness (the strict equality of the total number of crossings is fortuitous in this case). We may also refer to Fig.(15) in the appendix for visualization of τ and the underlying statistical t-test.

5 Conclusion

We propose a novel SSA criterion that emphasizes sign accuracy and zero-crossings of the predictor subject to a holding time constraint. Under the Gaussian assumption, the classic MSE-criterion is equivalent to unconstrained SSA-optimization: in the absence of a holding time constraint and down to an arbitrary scaling nuisance. We argue and demonstrate that the proposed concept is resilient against various departures from the Gaussian assumption. Moreover, the approach is interpretable and appealing due to its actual simplicity and because the criterion merges relevant concepts of prediction in terms of sign accuracy, MSE, and smoothing requirements, which are constituents of a prediction trilemma. In its primal form, the criterion aims at tracking the target optimally subject to an imposed noise suppression; in its dual form, the predictor generates the least zero-crossings for given track accuracy. An alternative to classic WH-smoothing can be obtained by inserting the identity for the target in the optimization criterion and the resulting SSA smoother addresses zero-crossings and MSE-performances explicitly. The business cycle case illustrates the application to a widely used benchmark, the HP filter, in view of modifying the classic predictor according to particular research priorities. In this context, timeliness and smoothness of real-time designs can be controlled effectively and interpretability or economic meaning can be maintained and preserved, due to optimality of the approximation. Our predictor outperforms the benchmark in a real-time BCA, across a selection of countries, in terms of fewer noisy crossings and smaller relative lead.

By applying SSA to a linear homoscedastic design, such as HP, our approach automatically retains the corresponding classic set of implicit or explicit assumptions. Therefore, topics such as non-linearity, asymmetry or heteroscedasticity are left as an avenue for future research. We conjecture that a combination of SSA with symmetric or asymmetric conditional heteroscedasticity models, such as SSA-GARCH or SSA-APARCH, could address some of these issues effectively.

6 Appendix

6.1 Solution to the SSA criterion

Let

$$M = \begin{pmatrix} 0 & 0.5 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 & 0 & 0 & 0 \\ \dots & & & & & & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0.5 & 0 \end{pmatrix}$$

of dimension $L \cdot L$ designate the so-called autocovariance-generating matrix so that $\rho(y, y, 1) = \frac{\mathbf{b}' \mathbf{M} \mathbf{b}}{\mathbf{b}' \mathbf{b}}$. Let λ_i, \mathbf{v}_i be the pairings of eigenvalues and (orthonormal) eigenvectors of \mathbf{M} , ordered according to the increasing size of λ_i , from smallest (negative) to largest (positive). Since the matrix \mathbf{V} of eigenvectors of \mathbf{M} defines an orthonormal basis of \mathbb{R}^L , the MSE estimate γ_δ can be expressed in the spectral representation

$$\gamma_\delta = \sum_{i=n}^m w_i \mathbf{v}_i = \mathbf{V} \mathbf{w} \quad (12)$$

with (spectral-) weights $\mathbf{w} = (w_1, \dots, w_L)'$ and where $1 \leq n \leq m \leq L$ and where $w_m \neq 0, w_n \neq 0$. We refer to γ_δ as having *complete* (or *incomplete*) spectral support depending on $w_i \neq 0$ for $i = 1, \dots, L$ (or not). The following theorem derives a generic form of the SSA solution: for simplicity, we here present shortened versions of the theorem as well as of its proof (complete theoretical results are added to the SSA-package).

Theorem 1. Consider the SSA optimization problem in Equation (5) or its extension to stationary autocorrelated processes in Section (2.3) with the following set of regularity assumptions:

1. $\gamma_\delta \neq 0$ (identifiability) and $L \geq 3$ (smoothing).
2. The SSA estimate \mathbf{b} is not proportional to γ_δ , denoted by $\mathbf{b} \not\propto \gamma_\delta$. In this case, the MSE estimate is not compliant with the holding time constraint and the latter cannot be dropped from the criterion (non-degenerate case).
3. $|\rho_1| < \rho_{\max}(L)$ (admissibility).
4. The MSE estimate γ_δ has complete spectral support, i.e., $w_i \neq 0$ for $i = 1, \dots, L$ (completeness).

If all regularity assumptions hold, then

1. the SSA predictor \mathbf{b} is given by

$$\mathbf{b} = \mathbf{b}(\nu) = \text{sign}^+ D \nu^{-1} \gamma_\delta \quad (13)$$

where $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, \dots, L\}$, $\nu := 2\mathbf{M} - \nu \mathbf{I}$ is an invertible $L \cdot L$ matrix, D is an arbitrary (positive) scaling term and $\text{sign}^+ = \pm 1$. Although b_{-1}, b_L do not explicitly appear in \mathbf{b} it is at least implicitly assumed that $b_{-1} = b_L = 0$ (implicit boundary constraints). The sign sign^+ is determined by requiring a positive criterion value and ν can be chosen such that \mathbf{b} complies with the holding time constraint.

2. The lag-one ACF $\rho(y(\nu), y(\nu), 1)$ and the target correlation $\rho(y(\nu), z, \delta)$ are strictly monotonic functions of ν , for $\nu > 2$, whereby $y(\nu)$ designates the output of the filter $\mathbf{b}(\nu)$ ¹².

Proof

For simplicity of exposition, we here discuss the SSA criterion (5) based on white noise $x_t = \epsilon_t$, acknowledging that straightforward modifications apply in the case of stationary autocorrelated processes x_t , see Section (2.3). We assume also that $\gamma_\delta' \gamma_\delta = 1$ which could always be achieved by arbitrary re-scaling since $\gamma_\delta \neq \mathbf{0}$ by the first regularity assumption (identifiability). Note that this re-scaling is not strictly necessary for a derivation of the proof but it simplifies notation. The SSA problem in Equation (5) can then be reformulated as

$$\begin{aligned} \max \gamma_\delta' \mathbf{b} \\ \mathbf{b}' \mathbf{b} &= 1 \\ \mathbf{b}' \mathbf{M} \mathbf{b} &= \rho_1 \end{aligned} \quad (14)$$

¹²A similar monotonicity argument applies for $\nu < -2$ but this case is less relevant in applications because the SSA predictor would generate more zero-crossings than the benchmark (smaller holding time).

where the additional constraint $\mathbf{b}'\mathbf{b} = 1$ (and the arbitrary scaling $\boldsymbol{\gamma}'_\delta \boldsymbol{\gamma}_\delta = 1$) ensures that covariances are also correlations, as in the original criterion. Consider the Lagrangian function

$$\mathcal{L} := \boldsymbol{\gamma}'_\delta \mathbf{b} - \lambda_1(\mathbf{b}'\mathbf{b} - 1) - \lambda_2(\mathbf{b}'\mathbf{M}\mathbf{b} - \rho_1) \quad (15)$$

Then the solution \mathbf{b} of the SSA problem must conform to the stationary Lagrangian or vanishing gradient equations

$$\boldsymbol{\gamma}_\delta = \lambda_1 2\mathbf{b} + \lambda_2(\mathbf{M} + \mathbf{M}')\mathbf{b} = \lambda_1 2\mathbf{b} + \lambda_2 2\mathbf{M}\mathbf{b} \quad (16)$$

Note that the second regularity assumption (non-degenerate case) implies that the holding time constraint in Equation (14) is 'active', i.e., $\lambda_2 \neq 0$. Dividing by λ_2 then leads to

$$D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}\mathbf{b} \quad (17)$$

$$\boldsymbol{\nu} := (2\mathbf{M} - \nu\mathbf{I}) \quad (18)$$

where $D = 1/\lambda_2$ and $\nu = -2\frac{\lambda_1}{\lambda_2}$. Equation (17) can be written as

$$\begin{aligned} b_{k+1} - \nu b_k + b_{k_1} &= D\gamma_{k+\delta}, \quad 1 \leq k \leq L-2 \\ b_1 - \nu b_0 &= D\gamma_\delta, \quad k=0 \\ -\nu b_{L-1} + b_{L-2} &= D\gamma_{L-1+\delta}, \quad k=L-1 \end{aligned} \quad (19)$$

for $k = 0, \dots, L-1$ so that $b_{-1} = b_L = 0$ are implicitly assumed for the natural extension $(b_{-1}, \mathbf{b}, b_L)'$ of the time-invariant linear filter. The eigenvalues of $\boldsymbol{\nu}$ are $2\lambda_i - \nu$ with corresponding eigenvectors \mathbf{v}_i . We note that if \mathbf{b} is the solution of the SSA problem, then $\nu/2$ cannot be an eigenvalue of \mathbf{M} since otherwise $\boldsymbol{\nu}$ in Equation (17) would map one of the eigenvectors in the spectral decomposition of \mathbf{b} to zero which would contradict the last regularity assumption (completeness). Therefore we can assume that $\boldsymbol{\nu}^{-1}$ exists, as claimed. A formal proof of the existence of $\nu = \nu(\rho_1)$, such that $\mathbf{b} = \mathbf{b}(\nu(\rho_1))$, defined by Equation (17), complies with the holding time constraint is skipped here. Briefly, a spectral decomposition of the lag-one autocorrelation determined by \mathbf{b} shows that the boundaries $\pm\rho_{max}(L)$ can be attained and an application of the intermediate-value theorem then implies that all $\rho_1 \in [-\rho_{max}(L), \rho_{max}(L)]$ can be matched by the lag-one ACF, as a function of ν . Finally, strict monotonicity of target correlation and lag-one ACF follow from inserting $\mathbf{b}(\nu)$ into the corresponding expressions, taking derivatives and showing that the latter neither vanish nor change the sign for $\nu > 2$ (the latter condition ensures that all eigenvalues of the matrix $\boldsymbol{\nu}$ are strictly negative). \square

Remarks

Gaussianity is not required in the derivation of the proof, because the SSA criterion (5) refers to correlations, solely. But Gaussianity can be invoked to link formally correlations with sign accuracy and holding time. The theorem derives exact finite-length solutions for any L such that $3 \leq L \leq T-1$. Therefore, the best predictor is always obtained for $L = T-1$ but the corresponding sample history would consist of a single point only, rendering direct comparisons with a benchmark impossible. If $sign^+ D = -\nu \rightarrow \infty$, then $sign^+ D\boldsymbol{\nu} \rightarrow \mathbf{I}$ so that $\mathbf{b}(\nu) \rightarrow \boldsymbol{\gamma}_\delta$. In this sense, the SSA-estimate $\mathbf{b}(\nu)$ generalizes the classic MSE predictor $\boldsymbol{\gamma}_\delta$. Finally, the condition $\nu > 2$, ensuring monotonicity in the second assertion of the theorem, is invariably obtained in a typical prediction framework, including a real-time BCA, assuming that $ht_1 > ht_{MSE}$ (smoothing¹³), where ht_{MSE} designates the holding time of the MSE predictor $\boldsymbol{\gamma}_\delta$.

¹³The case $\nu < -2$ covers situations such as exemplified by the rightmost column of Table (4), where SSA is asked to generate additional noisy crossings over the benchmark: $ht_1 < ht_{MSE}$. The SSA configuration with $-2 \leq \nu \leq 2$ can address very large (or very small) holding times ht_1 , requiring unit root filters, i.e., filters whose coefficients do not decay towards zero, due to the imposed smoothing constraint. For very large L such an extreme setting would eventually conflict with the stationarity assumption and it conflicts with an application to BCA, where ht_1 is calibrated to a benchmark.

Corollary 1. Let the assumptions of Theorem (1) hold. Then the solution to the SSA-optimization problem in Criterion (5) is

$$\mathbf{b}(\nu_0) = \text{sign}^+ \left(2\mathbf{M} - \nu_0 \mathbf{I} \right)^{-1} D\boldsymbol{\gamma}_\delta \quad (20)$$

where ν_0 is a solution to the non-linear equation

$$\frac{\mathbf{b}(\nu_0)' \mathbf{M} \mathbf{b}(\nu_0)}{\mathbf{b}(\nu_0)' \mathbf{b}(\nu_0)} = \rho_1 \quad (21)$$

The sign $\text{sign}^+ = \pm 1$ in Equation (20) is selected such that $\mathbf{b}(\nu_0)' \boldsymbol{\gamma}_\delta > 0$ (positive criterion value).

A proof follows readily from Theorem (1). We next derive the distribution of the SSA predictor.

Corollary 2. Let all regularity assumptions of Theorem (1) hold and let $\hat{\boldsymbol{\gamma}}_\delta$ be a finite-sample estimate of the MSE predictor $\boldsymbol{\gamma}_\delta$ with mean $\boldsymbol{\mu}_{\boldsymbol{\gamma}_\delta}$ and variance $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}_\delta}$. Then mean and variance of the SSA predictor $\hat{\mathbf{b}}$ are

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{b}} &= \text{sign}^+ D \boldsymbol{\nu}^{-1} \boldsymbol{\mu}_{\boldsymbol{\gamma}_\delta} \\ \boldsymbol{\Sigma}_{\mathbf{b}} &= D^2 \boldsymbol{\nu}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{\gamma}_\delta} \boldsymbol{\nu}^{-1} \end{aligned}$$

If $\hat{\boldsymbol{\gamma}}_\delta$ is Gaussian distributed then so is $\hat{\mathbf{b}}$.

The proof readily follows from Equation (13), noting that $\boldsymbol{\nu}^{-1}$ is symmetric so that $(\boldsymbol{\nu}^{-1})' = \boldsymbol{\nu}^{-1}$ in the expression for the variance. We here refer to standard textbooks for a derivation of mean, variance and (asymptotic) distribution of the MSE estimate under various assumptions about x_t , see Brockwell and Davis (1993). The last corollary proposes a dual derivation and interpretation of the SSA predictor.

Corollary 3. Let all regularity assumptions of Theorem (1) hold and let $y_t(\nu_0)$ denote the SSA solution in which $\nu_0 > 2$. Set $\rho_{\nu_0, \delta} := \rho(y(\nu_0), z, \delta) > 0$ and consider the dual optimization problem

$$\left. \begin{array}{l} \max_{\mathbf{b}} \rho(y, y, 1) \\ \rho(y, z, \delta) = \rho_{\nu_0, \delta} \end{array} \right\} \quad (22)$$

A solution to this problem has the same functional form $\mathbf{b} = \mathbf{b}(\nu)$ as in Equation (13) and if the search for ν can be restricted to the set $\{\nu | \nu > 2\}$, then $y_t(\nu_0)$ is also the solution to the dual problem.

Proof

The Lagrangian Equation (16) does not discern constraint and objective: after suitable re-scaling of Lagrange multipliers, the problem specified by Criterion (22) leads to the same functional form $\mathbf{b} = \text{sign}^+ D \boldsymbol{\nu}^{-1} \boldsymbol{\gamma}_\delta$ of its solution. The only difference to the original Criterion (5) is that ν in Criterion (22) must be selected such that $\rho(y(\nu), z, \delta) = \rho_{\nu_0, \delta}$. If the search can be restricted to $\{\nu | \nu > 2\}$, then the solution to the primal problem is also the solution to the dual problem, due to the strict monotonicity of $\rho(y(\nu), z, \delta)$. \square

In a typical prediction framework, including a real-time BCA, the restriction of the solution space to $\{\nu | \nu > 2\}$ in the above Corollary is not a limitation, see the final remark after the proof of Theorem (1). Corollary (3) then asserts that for a given tracking-ability, expressed in terms of $\rho_{\nu_0, \delta} > 0$, the SSA predictor generates the least number of zero-crossings or alarms, in the long-run. If the filter is optimized for data in first differences, as done in the BCA example below, then the transformed predictor for data in levels is maximally monotonic because sign changes of its drift are the fewest possible.

6.2 Extension to Non-Stationary Integrated Processes

We here propose an extension of Section (2.3). If x_t is a non-stationary integrated process, then we require the target filter γ_k to cancel its unit root(s) so that \tilde{z}_t is stationary; otherwise, the concept of zero-crossings would not be meaningfully defined anymore. For illustration, assume that x_t has a single unit root at frequency zero so that $\Delta x_t := x_t - x_{t-1} = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k}$, where $\xi_k, k \geq 0$ is a square-summable sequence. In this case, we wish to replicate \tilde{z}_t by a filter $\tilde{\gamma}_k$ such that

$$\tilde{z}_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k} = \sum_{k=-\infty}^{\infty} \tilde{\gamma}_k \Delta x_{t-k} = \sum_{k=-\infty}^{\infty} (\tilde{\gamma} \cdot \xi)_k \epsilon_{t-k} \quad (23)$$

where we assume that the non-stationary process x_t has been initialized sometime, $x_{t_0} = x_0$ for some $t_0 \in \mathbb{Z}$ and $x_0 \in \mathbb{R}$. We infer that $\tilde{\gamma}_k = \sum_{j=-\infty}^k \gamma_j$ is the convolution of γ_k and of the one-sided integration or summation filter, denoted by $\tilde{\gamma} = \Delta^{-1}\gamma$. If the filter with weights $\tilde{\gamma}_k$ is well-defined, then the above equality holds, the non-stationarity is resolved and the extension proposed in Equation (6) applies, substituting $\tilde{\gamma}_k$ for γ_k . The above transformation can be generalized to arbitrary integration orders d of x_t , by applying differences of order d to x_t , $\Delta^d x_t = \tilde{x}_t$, and sum-convolutions of order d to γ_k , $\tilde{\gamma} = \Delta^{-d}\gamma$. We now seek for conditions such that $\sum_{k=-\infty}^{\infty} (\tilde{\gamma} \cdot \xi)_k^2 < \infty$ in (23), so that \tilde{z}_t is stationary.

Proposition 5. *Let x_t be an integrated processes with a unit root of order d at frequency zero and assume the spectral density $h_{\tilde{x}}(\omega) = |\sum_{k=0}^{\infty} \xi_k \exp(-ik\omega)|^2 / (2\pi)$ of the stationary process $\tilde{x}_t := \Delta^d x_t = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k}$ exists for $\omega \in [-\pi, \pi]$, whereby ϵ_t is standardized white noise. If γ_k is such that $\sum_{k=-\infty}^{\infty} |\gamma_k k^d| < \infty$ and $\sum_{k=-\infty}^{\infty} \gamma_k k^j = 0$, for $j = 0, \dots, d-1$, then \tilde{z}_t is a stationary process with spectral density $h_{\tilde{z}} = |\tilde{\Gamma}(\omega)|^2 h_{\tilde{x}}(\omega)$, where $\tilde{\Gamma}(\omega)$ denotes the transfer function of the filter with weights $\Delta^{-d}\gamma$.*

Proof

For the sake of simplicity, we assume $d = 1$, indicating where necessary modifications of the proof in the case $d > 1$. Let us first look at the convolution of filters on the right side of Equation (23), after z-transformation in the frequency domain (convolution theorem):

$$\sum_{k=-\infty}^{\infty} \tilde{\gamma}_k \exp(-ik\omega) \sum_{j=0}^{\infty} \xi_j \exp(-ij\omega) = \frac{\sum_{k=-\infty}^{\infty} \gamma_k \exp(-ik\omega)}{1 - \exp(-i\omega)} \sum_{j=0}^{\infty} \xi_j \exp(-ij\omega)$$

Since $\sum_{k=-\infty}^{\infty} |\gamma_k k| < \infty$, by assumption, we infer that the term on the right-hand side is well defined for all $\omega \neq 0$ and that the derivative of the numerator with respect to ω exists everywhere. Relying on first-order Taylor approximations of numerator and denominator, centered at $\omega = 0$ and using $\sum_{k=-\infty}^{\infty} \gamma_k = 0$, we then obtain

$$\lim_{\omega \rightarrow 0} \frac{\sum_{k=-\infty}^{\infty} \gamma_k \exp(-ik\omega)}{1 - \exp(-i\omega)} = \frac{\sum_{k=-\infty}^{\infty} -ik\gamma_k}{i} = - \sum_{k=-\infty}^{\infty} k\gamma_k$$

so that $\tilde{\Gamma}(\omega) := \frac{\sum_{k=-\infty}^{\infty} \gamma_k \exp(-ik\omega)}{1 - \exp(-i\omega)}$, with weights $\tilde{\gamma}_k$, is well defined, in terms of the above limit, for all $\omega \in [-\pi, \pi]$. If $d > 1$, then a similar proof applies by relying on d -th order Taylor approximations, using $\sum_{k=-\infty}^{\infty} \gamma_k k^j = 0$, $j = 0, \dots, d-1$, for canceling the unit root singularity. From the above, we then infer that \tilde{z}_t is a stationary process with spectral density $h_{\tilde{z}}(\omega) = |\tilde{\Gamma}(\omega)|^2 h_{\tilde{x}}(\omega)$. \square

Remarks

Well-known examples of unit root cancelling filters are two- and one-sided HP gap filters, $x_t - \tilde{z}_{t,HP}$ and $x_t - y_{t,HP}$, two- and one-sided BK bandpass and the (one-sided) Hamilton filter: the two-sided HP gap can cancel unit roots up to order $d = 4$ ($d = 2$ for the one-sided gap); the other

filters can accommodate higher integration orders, when implemented accordingly. Also, the above proposition is useful when evaluating business cycle indicators because $|\tilde{\Gamma}(\omega)|^2 h_{\tilde{x}}(\omega)$ measures the spectral content of the extracted cycle \tilde{z}_t . It is then commode to rely on classic frequency-domain analysis to address smoothing capability, leads or lags of competing predictors (nowcasting) or more specific problems, like the generation of ‘spurious’ cycles by a filter.

6.3 Finite Sample Estimation Error

We briefly inspect the effect of the finite-sample estimation error on the holding time of the SSA predictor. For illustration, we rely on a simulation experiment based on the HP trend filter applied to AR(1) processes $x_t = a_1 x_{t-1} + \epsilon_t$, for various specifications of a_1 covering a wide range of use-cases and assuming a sample length $T = 120$ corresponding to ten years of monthly data. For each process, we compute 1000 realizations of a SSA(12,0) nowcast by relying on the generalized Criterion (6), inserting $\xi_j = \hat{a}_1^j$ in the target correlation as well as in the lag-one ACF (7), whereby \hat{a}_1 is the classic (MSE) estimate of a_1 (we here rely on the R-function arima for estimation). Out-of-sample holding times are then obtained by inserting the true lag-one ACF of each empirical nowcast into Equation (3), where the ACF is based on Equation (7), using $\xi_j := a_1^j$, the true DGP ($\xi_j := \hat{a}_1^j$ would lead to the fixed in-sample value $ht_1 = 12$). Means and standard deviations of out-of-sample holding times are reported in Table (12) together with HP concurrent, for reference. The table suggests that SSA maintains a firm grip on the holding time in the sense that standard

	AR(1)=-0.9	-0.5	-0.2	0	0.2	0.5	0.9
HP ht	2.15	4.58	6.32	7.66	9.28	12.84	28.63
SSA(12,0) mean ht	11.92	11.93	11.96	11.95	11.98	11.92	12.93
SSA(12,0) sd ht	0.85	0.46	0.49	0.59	0.74	1.24	3.34

Table 12: Means and standard deviations of out-of-sample holding times of SSA(12,0) based on samples of length 120 (corresponding to 10 years of monthly data). Holding times of the classic concurrent HP are added for reference.

deviations are small, in relative terms, and that biases are negligible, except for strongly autocorrelated processes in the rightmost column. In the latter case, the MSE estimate \hat{a}_1 is negatively biased and left-skewed, see Hamilton (1994), and SSA does not sufficiently compensate for the strong dependence of the data so that the effective (out-of-sample) lag-one ACF and holding time are positively biased (a similar but reversed and less pronounced effect can be observed for small a_1 , too). The increase in variance for larger a_1 is due to the non-linear transformation in Equation (3) which stretches positive outliers of the lag-one ACF. However, these undesirable effects are mostly irrelevant because the dependence structure of typical economic data is generally weak, at least after applying first differences. Moreover, one of the main purposes of our approach consists in providing an explicit control of zero-crossings when referenced against a benchmark so that the ‘true’ or absolute holding time, emphasized in the above simulation experiment, is to some extent less relevant than the relative holding time *difference* (or ratio) between SSA and the benchmark. Since the sampling error $a_1 - \hat{a}_1$ operates similarly on benchmark and SSA, its relative effect is subject to cancellation in this case.

6.4 Measuring Advancement and Retardation at Zero-Crossings: τ -Statistic

To quantify leads or lags of filters at zero-crossings we here propose a simple formal test statistic. Let y_{tn} , $n = 1, \dots, N$ be a set of competing filters and let ZC_n denote the set of zero-crossings $t_{jn}, j = 1, \dots, |ZC_n|$ of each filter y_{tn}

$$ZC_n = \{t_{jn} | \text{sign}(y_{t_{jn},n}) \neq \text{sign}(y_{t_{j-1,n}})\}$$

where $|ZC_n|$ means the cardinality of the set. Let ZC_n^+ and ZC_n^- designate the sub-sets of up- and downturns of y_{tn} , at which y_{tn} crosses the zero-line from below or from above. Assume that y_{tN} has been selected to be benchmarked against filter $n < N$ according to the following measure

$$\tau(N, n) := \frac{1}{|ZC_N|} \left(\sum_{j=1}^{|ZC_N^+|} (t_{f(j,N),n}^+ - t_{jN}^+) + \sum_{j=1}^{|ZC_N^-|} (t_{f(j,N),n}^- - t_{jN}^-) \right) \quad (24)$$

where $f(j, N)$ is the index of the zero-crossing of y_{tn} closest to t_{jN}^+ or t_{jN}^- in the corresponding subsets ZC_n^+ or ZC_n^- , i.e., $|t_{f(j,N),n}^+ - t_{jN}^+| = \min_i |t_{in}^+ - t_{jN}^+|$ and similarly for the downturns. Note that sums are taken over zero-crossings of the reference filter y_{tN} , which is always an SSA design in the empirical sections. We recommend that y_{tn} , $n = 1, \dots, N$ should have ‘similar’ crossings for the comparison to be meaningful (comparing a lowpass to a highpass would lead to difficulties when interpreting results); also, ideally, the reference filter should be smoother, with fewer zero-crossings, as is the case in our examples. In our BCA application, we mainly rely on SSA(12,18): we prefer this reference to the proper target, i.e., the HP symmetric filter, because the latter cannot be used towards the sample-end, thus excluding the latest and important great-lockdown crisis, see Fig.(14). Also, timeliness performances of all concurrent filters can be assessed against another causal filter, which facilitates direct comparisons.

The τ -statistic (24) is called *mean shift* of the reference filter N with respect to filter n ; the former is called *leading* or *lagging*, with respect to the latter, depending on the mean-shift being positive or negative. The statistic could be split into separate downturn and upturn sums in the case of asymmetry. Classic t-tests

$$\begin{aligned} t^+ &= \frac{\text{Mean}(t_{f(j,N),n}^+ - t_{jN}^+)}{\text{Sd}((t_{f(j,N),n}^+ - t_{jN}^+))} \\ t^- &= \frac{\text{Mean}(t_{f(j,N),n}^- - t_{jN}^-)}{\text{Sd}((t_{f(j,N),n}^- - t_{jN}^-))} \end{aligned}$$

where $\text{SD}()$ corresponds to the empirical standard deviation, can be used to infer statistical significance of a mean-lead or a -lag at up- and downturns, assuming the summands in equation (24) to be independently distributed. Note that one could merge both statistics into a single one, when assuming symmetry: we here rely on the latter single number because timing *differences* of competing designs, as measured by the τ -statistics, are to some extent less sensitive to unequal lengths of positive and negative half-cycles. For illustration, Fig.(15) displays the cumulated shifts at zero-crossings of the filters in Table (11) (single long concatenation of all country-specific series).

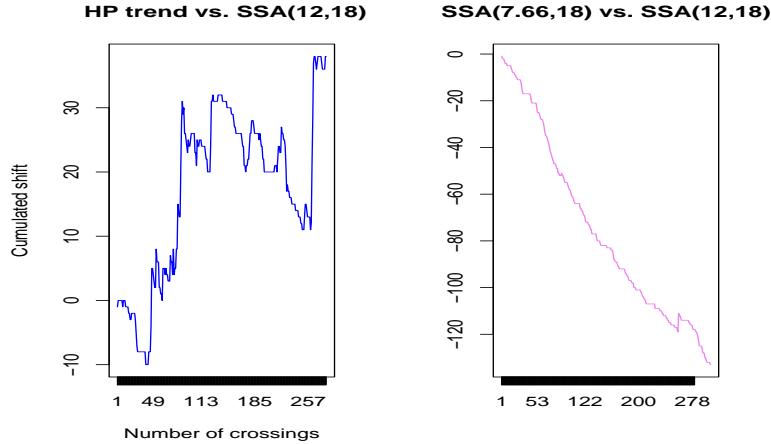


Figure 15: Cumulated Shift at zero-crossings computed on concatenated series: HP trend (left panel) and SSA(7.66,18) (right panel) are both referenced against SSA(12,18). An upward trend signifies a relative lead of the reference SSA(12,18).

The mean-shift or τ -statistic corresponds to the slope of these curves and a positive slope signifies a lead of the reference design. The (merged) t-statistic tests the alternative (non-vanishing drift) against the null (vanishing drift): the large absolute t-value for the second SSA filter in Table (11) reflects the strong drift in the right panel of Figure (15).

References

- [1] Anderson O.D. (1975) Moving Average Processes. *Journal of the Royal Statistical Society. Series D (The Statistician)*. **Vol. 24, No. 4**, 283-297
- [2] Barnett J.T. (1996) Zero-crossing rates of some non-Gaussian processes with application to detection and estimation. *Thesis report Ph.D.96-10, University of Maryland*.
- [3] Baxter, M. and R.G. King (1999). Measuring business cycles: Approximate band-pass filters for economic time series. *Review of Economics and Statistics* **81**, 575–593.
- [4] Brockwell P.J. and Davis R.A. (1993) Time Series: Theories and Methods (second edition). *Springer Verlag*.
- [5] Cornea-Madeira, Adriana (2017). The Explicit Formula for the Hodrick-Prescott Filter in Finite Sample. *Review of economics and statistics*, 314-318.
- [6] Davies, N., Pate, M. B. and Frost, M. G. (1974). Maximum autocorrelations for moving average processes. *Biometrika* **61**, 199-200.
- [7] Fama, Eugene F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business* **38(1)**, 34-105.
- [8] Gómez, V. (2001). The use of Butterworth filters for trend and cycle estimation in economic time series. *Journal of Business and Economic Statistics* **19**, 365–373.
- [9] Granger, C.W.J. (1966). The typical spectral shape of an economic variable. *Econometrica* **34**, 150-161.
- [10] Hall, R. E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence *Journal of Political Economy* **6(6)**, 971-987.

- [11] Hamilton, J.D. (1994). Time Series Analysis. *Princeton University Press*.
- [12] Hamilton, J.D. (2009). Understanding Crude Oil Prices *Energy Journal* **30**(2), 179- 206.
- [13] Hamilton, J.D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics* **100**, 831–843.
- [14] Harvey, A. 1989. Forecasting, structural time series models and the Kalman filter. *Cambridge: Cambridge University Press*.
- [15] Hodrick, R. and Prescott, E. (1997) Postwar U.S. business cycles: an empirical investigation. *Journal of Money, Credit, and Banking* **29**, 1–16.
- [16] Kedem, B. (1986) Zero-crossings analysis. *Research report AFOSR-TR-86-0413, Univ. of Maryland*.
- [17] Kratz, M. (2006) Level crossings and other level functionals of stationary Gaussian processes. *Probability surveys* **Vol. 3**, 230-288.
- [18] Mankiw, N. G. (1987). The Optimal Collection of Seigniorage: Theory and Evidence. *Journal of Monetary Economics* **20**(2), 327-341.
- [19] McElroy, T. (2006) Exact Formulas for the Hodrick-Prescott Filter. *Research report series (Statistics 2006-9). U.S. Census Bureau* .
- [20] McElroy, T. and Wildi , M. (2019) The trilemma between accuracy, timeliness and smoothness in real-time signal extraction. *International Journal of Forecasting* **35** (3), 1072-1084.
- [21] McElroy, T. and Wildi , M. (2020) The multivariate linear prediction problem: model-based and direct filtering solutions. *Econometrics and Statistics* **14**, 112-130.
- [22] Morten, O. and Uhlig, H. (2002) On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *The Review of Economics and Statistics* **84** (2), 371-376.
- [23] Osterrieder, J. (2017) The Statistics of Bitcoin and Cryptocurrencies. *Advances in Economics, Business and Management Research (AEBMR)* **Vol. 26**.
- [24] Phillips, P.C.B., S. Jin (2021). Business cycles, trend elimination, and the HP filter. *International Economic Review* **62**, 469–520.
- [25] Phillips, P.C.B., Z. Shi (2021). Boosting: Why you can use the HP filter. *International Economic Review* **62**, 521–570.
- [26] Rice,S.O. (1944). Mathematical analysis of random noise. *I. Bell. Syst. Tech. J* **23**, 282-332.
- [27] Samuelson, P. (1965). Proof that Properly Anticipated Prices Fluctuate Randomly.*Industrial Management Review* **6**(2), 41-49.
- [28] Sargent, T.J. (1976). A Classical Macroeconometric Model for the United States. *Journal of Political Economy* **84**(2), 207-237.
- [29] Whittaker, E.T. (1922). On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society* , **41** , 63 - 75.
- [30] Wildi, M. (2005). Signal Extraction. *Lecture Notes in Economics and Mathematical Systems*, Springer.