Sign Accuracy, Mean-Square Error and Zero Crossings: a Generalized Forecast Approach

Marc Wildi

June 10, 2024

Abstract

We propose an extension of classic time series forecast approaches 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 a time series mark transitions between expansion and contraction episodes which can be relevant for decision-making and control. 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. In this context, we propose and illustrate a 'customization' of existing benchmark predictors in terms of ATS performances. Finally, we demonstrate that a simple special case of SSA can address smoothing and we benchmark performances against an established benchmark.

1 Introduction

Forecasting is the process of making 'optimal' predictions based on current and past realizations of a (set of) time series. Typically, optimality refers to prediction accuracy or, more precisely, to the minimization of the forecast error between a future target and the predictor. While this proceeding seems uncontroversial, in principle, we argue that alternative characteristics of a predictor might draw attention such as the smoothing capability, i.e., the extent by which undesirable 'noisy' components can be suppressed, or timeliness, as measured by advancement (lead, left-shift) and retardation (lag, right-shift), or sign accuracy and zero-crossings, as measured by the ability to predict the sign of the target. Wildi (2024) proposes a generic forecast approach, called Simple Sign Accuracy (SSA), which merges sign accuracy and mean-square error (MSE) performances subject to a holding time constraint that determines the duration between consecutive zero-crossings of the predictor. We here generalize the approach to non-stationary integrated processes, we derive a novel closed-form solution of the optimization problem and we provide a rigorous theoretical underpinning as well as an exhaustive illustration of the main traits and capabilities of SSA in the time- as well as in the frequency domain.

Wildi (2024) investigates business-cycle analysis (BCA) and proposes an application of SSA to trend-nowcasting which represents a challenging prediction problem, encompassing classic one- and multi-steap ahead forecasting. Ideally, accuracy, timeliness and smoothness (ATS) of a (real-time) BCA indicator can be addressed and optimized simultaneously, by a fixed optimization criterion, such that the resulting nowcast delivers the closest approximation to the true but unobserved level of the trend (Accuracy), while tracking sign-changes immediately, without systematic delay (Timeliness), and eventually avoiding the generation of 'false' crossings (Smoothness). 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 tradeoffs 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. SSA allows users to control and balance all three aspects of the trilemma to take account of zero-crossings, either directly or by expanding existing popular approaches (benchmark customization).

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* (HT), 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. SSA as introduced in Wildi (2024) controls explicitly for the HT of a predictor 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 SSA. Finally, we demonstrate that our approach allows for a straightforward customization of benchmark predictors in terms of ATS performances.
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-package[1] which extends the application field to BCA, including SSA-customizations of the HP-filter, of Hamilton's regression filter, Hamilton (2018), and of the Baxter-King (BK) bandpass filter, Baxter and King (1999).

Section (2) introduces the SSA criterion based on a simplified framework; a technical treatment is proposed in Section (3) with a derivation of numerical and closed-form solutions in regular and singular cases; Section (4) explains and illustrates some of the distinguishing features of the predictor and Section (5) proposes applications to time series forecasting and real-time signal extraction, highlighting a prediction trilemma whose constituents can be controlled by SSA; an extension of the basic framework is proposed in Section (6) and Section (7) addresses smoothing; finally, Section (8) summarizes our main findings.

[1] An 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.

2 Simple Sign-Accuracy (SSA-) Criterion

We here briefly introduce the SSA-criterion proposed in Wildi (2024) and provide additional context. We assume a zero-mean stationary time series x_t (data) and a target $z_{t+\delta}$, $\delta \in \mathbb{Z}$, which depends on future x_{t-k}, $k < 0$. We then derive an optimal predictor y_t of $z_{t+\delta}$ based on x_{t-k}, $k \geq 0$ (causal filter) with the property that the HT of y_t can be specified by the user. For illustration, and for simplicity of exposition, we first assume $x_t = \epsilon_t$ to be a white noise (WN) sequence: extensions to autocorrelated stationary and non-stationary integrated processes do not affect the main theoretical results and are discussed in Section (6). Also, the HT is formulated in terms of the autocorrelation function (ACF) of the predictor and a formal link between the HT and the ACF is established later, in Section (4).

Let then $z_t = \sum_{k=-\infty}^{\infty} \gamma_k \epsilon_{t-k}$, where $\epsilon_j, j \in \mathbb{Z}$, is white noise (for simplicity of exposition we may assume ϵ_t to be standardized), and $\boldsymbol{\gamma} := (\gamma_k), k \in \mathbb{Z}$, is a (real) square summable sequence so that z_t is a stationary zero-mean process with variance $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We look for a predictor $y_t := \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$ of the target $z_{t+\delta}$, $\delta \in \mathbb{Z}$, 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. Classic one- or $h-$step ahead forecasting is obtained by setting $z_t = x_t$ and $\delta = h$ but our framework can accommodate for more complex targets such as, e.g., trend or cycle components. We here restrict attention to univariate forecast problems, where the target and the predictor depend on a single series: an extension to multivariate problems must be deferred, due to length constraints. Consider now the following optimization problem

$$\left.\begin{array}{c} \max_{\mathbf{b}} \mathbf{b}'\boldsymbol{\gamma}_\delta \\ \mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1 \\ \mathbf{b}'\mathbf{b} = l \end{array}\right\}, \tag{1}$$

where $\mathbf{b} = (b_0, ..., b_{L-1})$, $\boldsymbol{\gamma}_\delta = (\gamma_\delta, ..., \gamma_{\delta+L-1})'$ are L-dim. column vectors, l is a scaling (constant) and

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

is of dimension $L \cdot L$ such that $\mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{k=1}^{L-1} b_{k-1} b_k$ is the lag-one autocovariance of y_t under the WN assumption. Criterion (1) is referred to as *simple sign accuracy* or SSA criterion and its solution is denoted by SSA(ρ_1, δ) (its dependence on l is omitted in the notation, see below for details); the constraints $\mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1$ and $\mathbf{b}'\mathbf{b} = l$ are referred to as HT and length constraints, respectively, see Wildi (2024). Under the WN assumption, the classic mean-square error (MSE) predictor is $\mathbf{b} := \boldsymbol{\gamma}_\delta$ and it can be obtained as a solution to the SSA criterion by setting $l := \boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta$ and $\rho_1 := \boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\gamma}_\delta / \boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta$.

To simplify terminology, we now merge the concepts of filter outputs and filter weights such that, e.g., $y_{t,MSE} := \boldsymbol{\gamma}_\delta'\boldsymbol{\epsilon}_t$ or $\boldsymbol{\gamma}_\delta$ will both be referred to as MSE predictor and similarly $y_t = \mathbf{b}'\boldsymbol{\epsilon}_t$ or $\mathbf{b}$ are the SSA predictor, where $\boldsymbol{\epsilon}_t := (\epsilon_t, ..., \epsilon_{t-(L-1)})'$. Under the WN assumption, the MSE-predictor $\boldsymbol{\gamma}_\delta$ stands as representative for the proper target $\boldsymbol{\gamma}$ in Criterion (1) in the sense that γ_k are irrelevant for $k < \delta$ or $k > \delta + L - 1$ and we now assume $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$. In this context, the solution $\mathbf{b}_0$ to the SSA criterion (or $y_{0t} := \mathbf{b}_0'\boldsymbol{\epsilon}_t$) can be interpreted as a (constrained) predictor for $z_{t+\delta}$, see Section (5) for context; alternatively, $\mathbf{b}_0$ can be viewed as a 'smoother' for $y_{t,MSE}$, see Section (7). In this sense, the SSA criterion merges prediction and smoothing. Also, $\mathbf{b}'\mathbf{M}\mathbf{b}/l = \mathbf{b}'\mathbf{M}\mathbf{b}/\mathbf{b}'\mathbf{b} =: \rho(y, y, 1)$ is the lag-one autocorrelation (ACF) of y_t and the objective function $\mathbf{b}'\boldsymbol{\gamma}_\delta$ is proportional to $\rho(y, z, \delta) := \mathbf{b}'\boldsymbol{\gamma}_\delta / \sqrt{l\boldsymbol{\gamma}'\boldsymbol{\gamma}}$, the target correlation of y_t with $z_{t+\delta}$, or to $\mathbf{b}'\boldsymbol{\gamma}_\delta / \sqrt{l\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, the correlation of y_t with $y_{t,MSE}$: maximizing either of these objective functions

maximizes the other ones too and therefore Criterion (1) is equivalent to

$$\left.\begin{array}{c}\max_{\mathbf{b}} \rho(y,z,\delta) \\ \rho(y,y,1)=\rho_1 \\ \mathbf{b}'\mathbf{b}=l\end{array}\right\}. \tag{2}$$

Intuitively, an increase of ρ_1 would lead to a stronger lag-one ACF and therefore to a 'smoother' path of y_t with less frequent zero-crossings, see Section (4) for context. If the HT-constraint is omitted, then the solution to the SSA criterion is $\sqrt{l}\boldsymbol{\gamma}_\delta/\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, the MSE predictor up to an arbitrary scaling. Since correlations or signs or zero-crossings are indifferent to the scaling of y_t, we consider the latter as a nuisance parameter, i.e., a quantity which is not of immediate interest for the analysis, and the length constraint merely ensures uniqueness: if explicitly required, an optimal scaling can always be obtained afterward, once the solution $\mathbf{b}=\mathbf{b}(l)$ has been computed for an arbitrary l. We now derive a solution to Criterion (1), acknowledging that our results will be extended to autocorrelated processes x_t.

3 Solution to the SSA-Criterion

Wildi (2024) sketches the SSA-solution for the so-called regular case. We here propose a rigorous treatment, addressing regular as well as singular cases. For this purpose, consider the (Fourier) eigenvectors $\mathbf{v}_j := \left(\sin(k\omega_j)/\sqrt{\sum_{k=1}^{L} sin(k\omega_j)^2}\right)_{k=1,...,L}$ of $\mathbf{M}$ with adjoined eigenvalues $\lambda_j = \cos(\omega_j)$ computed at the discrete Fourier frequencies $\omega_j = j\pi/(L+1)$, $j=1,...,L$, see Anderson (1975): we normalize the eigenvectors to constitute an orthonormal basis of $\mathbb{R}^L$, which will simplify subsequent notation.

Proposition 1. *Under the above assumptions, the vector* $\mathbf{b}$ *is a stationary point of the lag-one ACF* $\rho(y,y,1)$ *if and only if* $\mathbf{b}$ *is an eigenvector* $\mathbf{v}_i$ *of* $\mathbf{M}$ *with corresponding eigenvalue* $\lambda_i=\rho(y,y,1)$*, for some* $i\in\{1,...,L\}$*. Furthermore, the lag-one ACF of a MA-filter of length* L *is bounded by* $\lambda_L=-\cos(\pi/(L+1))=\rho_{min}(L)\leq\rho(y,y,1)\leq\rho_{max}(L)=\cos(\pi/(L+1))=\lambda_1$*. Maximum and minimum ACF values are achieved for* $\mathbf{b}:=\sqrt{l}\mathbf{v}_1$ *and* $\mathbf{b}:=\sqrt{l}\mathbf{v}_L$*, respectively.*

Proof: Assume, for simplicity of exposition, that $\mathbf{b}'\mathbf{b}=l=1$ so that $\rho(y,y,1)=\mathbf{b}'\mathbf{M}\mathbf{b}$. A stationary point of $\rho(y,y,1)$ is found by equating the derivative of the Lagrangian $\mathfrak{L}=\mathbf{b}'\mathbf{M}\mathbf{b}-\lambda(\mathbf{b}'\mathbf{b}-1)$ to zero i.e. $(\mathbf{M}+\mathbf{M}')\mathbf{b}=2\mathbf{M}\mathbf{b}=2\lambda\mathbf{b}$. We deduce that $\mathbf{b}$ is a stationary point if and only if it is an eigenvector of $\mathbf{M}$. Then $\rho(y,y,1)=\mathbf{b}'\mathbf{M}\mathbf{b}=\lambda_i\mathbf{b}'\mathbf{b}=\lambda_i$ for some $i\in\{1,...,L\}$ and therefore $\rho(y,y,1)$ must be the corresponding eigenvalue, as claimed. Since the unit-sphere is free of boundary-points, we conclude that the extremal values $\rho_{min}(L)$, $\rho_{max}(L)$ must be stationary points so that $\rho_{min}(L)=-\cos(\pi/(L+1))=\lambda_L$ and $\rho_{max}(L)=\cos(\pi/(L+1))=\lambda_1$ and the boundary values are reached by $\mathbf{b}:=\mathbf{v}_L$ and $\mathbf{b}:=\mathbf{v}_1$, respectively, as claimed. □

We now introduce the spectral decomposition of the MSE-filter $\boldsymbol{\gamma}_\delta\neq\mathbf{0}$:

$$\boldsymbol{\gamma}_\delta=\sum_{i=n}^{m} w_i\mathbf{v}_i=\mathbf{V}\mathbf{w} \tag{3}$$

with (spectral-) weights $\mathbf{w}=(w_1,...,w_L)'$, where $1\leq n\leq m\leq L$ and $w_m\neq 0, w_n\neq 0$. If $n>1$ or $m<L$ then the MSE predictor $\boldsymbol{\gamma}_\delta$ is called *band-limited*. Also, we refer to $\boldsymbol{\gamma}_\delta$ as having either *complete* or *incomplete* spectral support depending on $w_i\neq 0$ for $i=1,...,L$ or not. Finally, denote by $NZ:=\{i|w_i\neq 0\}$ the set of indexes of non-vanishing weights w_i so that $NZ=\{1,2,...,L\}$ iff $\boldsymbol{\gamma}_\delta$ has complete spectral support in which case it is not band-limited.

Corollary 1. *Consider the SSA Criterion* (1)*. If* $\rho_1<\lambda_L$ *or* $\rho_1>\lambda_1$ *then the problem does not admit a solution. If* $\rho_1=\lambda_1$ *or* $\rho_1=\lambda_L$ *and if* $\boldsymbol{\gamma}_\delta$ *is not band-limited, then the SSA solutions are* $\mathbf{b}_1:=sign(w_1)\sqrt{l}\mathbf{v}_1$ *and* $\mathbf{b}_L:=sign(w_L)\sqrt{l}\mathbf{v}_L$*, respectively.*

A proof follows directly from Proposition (1), noting that $\mathbf{b}_1'\boldsymbol{\gamma}_\delta = \text{sign}(w_1)\sqrt{l}w_1 > 0$ and $\mathbf{b}_L'\boldsymbol{\gamma}_\delta = \text{sign}(w_L)\sqrt{l}w_L > 0$, where strict positiveness applies due to maximization[2] and because $\boldsymbol{\gamma}_\delta$ is not band-limited.

We now derive the solution to the SSA criterion, assuming $\boldsymbol{\gamma}_\delta$ to have complete spectral support, which is referred to as the *regular* case.

Theorem 1. *Consider the SSA Criterion* (1) *and assume that the following set of regularity assumptions hold:*

1. $\boldsymbol{\gamma}_\delta \neq 0$ *(identifiability) and* $L \geq 3$*;*
2. *the SSA estimate* $\mathbf{b}$ *is not proportional to* $\boldsymbol{\gamma}_\delta$*, denoted by* $\mathbf{b} \not\propto \boldsymbol{\gamma}_\delta$ *(non-degenerate case);*
3. $|\rho_1| < \rho_{max}(L)$ *(admissibility);*
4. *the MSE-estimate* $\boldsymbol{\gamma}_\delta$ *has complete spectral support (completeness).*

Then:

1. *The solution to Criterion* (1) *has the one-parametric form*

$$\mathbf{b}(\nu) = D(\nu, l)\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta = D(\nu, l)\sum_{i=1}^{L}\frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i, \tag{4}$$

   where $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, ..., L\}$*,* $D = D(\nu, l) \neq 0$ *and* $\mathbf{N} := 2\mathbf{M} - \nu\mathbf{I}$ *is an invertible* $L \cdot L$ *matrix. Although* $b_{-1}(\nu), b_L(\nu)$ *do not explicitly appear in* $\mathbf{b}(\nu)$ *it is at least implicitly assumed that* $b_{-1}(\nu) = b_L(\nu) = 0$ *(implicit boundary constraints). Also,* $D(\nu, l)$ *is determined by* ν *and the length constraint; in particular, its sign is determined by asking for a positive objective function.*

2. *The lag-one ACF of* $y_t(\nu)$*, where* $y_t(\nu)$ *denotes the output of* $\mathbf{b}(\nu)$*, is*

$$\rho(\nu) := \rho(y(\nu), y(\nu), 1) = \frac{\mathbf{b}(\nu)'\mathbf{M}\mathbf{b}(\nu)}{\mathbf{b}(\nu)'\mathbf{b}(\nu)} = \frac{\sum_{i=1}^{L}\lambda_i w_i^2 \frac{1}{(2\lambda_i - \nu)^2}}{\sum_{i=1}^{L} w_i^2 \frac{1}{(2\lambda_i - \nu)^2}}. \tag{5}$$

   Moreover, $\nu = \nu(\rho_1)$ *can always be found such that* $y_t(\nu(\rho_1))$ *complies with the HT constraint.*

3. *The derivative* $d\rho(\nu)/d\nu$ *is strictly negative for* $\nu \in \{x | |x| > 2\rho_{max}(L)\}$*. Moreover,*

$$max_{\nu < -2\rho_{max}(L)}\rho(\nu) = min_{\nu > 2\rho_{max}(L)}\rho(\nu) = \rho_{MSE},$$

   where ρ_{MSE} *denotes the lag-one ACF of* $y_{t,MSE}$*.*

4. *For* $\nu \in \{x | |x| > 2\rho_{max}(L)\}$ *the derivatives of the objective function and lag-one ACF (with respect to* ν*) are linked by*

$$\frac{d\rho(y(\nu), z, \delta)}{d\nu} = -sign(\nu)\frac{\sqrt{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta}}{\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\frac{d\rho(\nu)}{d\nu} < 0. \tag{6}$$

Proof: Define the Lagrangian $\mathcal{L} := \boldsymbol{\gamma}_\delta'\mathbf{b} - \tilde{\lambda}_1(\mathbf{b}'\mathbf{b} - 1) - \tilde{\lambda}_2(\mathbf{b}'\mathbf{M}\mathbf{b} - \rho_1)$, where we assume $l = 1$ in Criterion (1) and where the 'tilde' distinguishes Lagrangian parameters from eigenvalues of $\mathbf{M}$. By assumption, $L \geq 3$ so that $\mathbf{b}$ is defined on a $L - 2 \geq 1$ dimensional intersection of unit-sphere

[2]The objective function is nearly overruled by the HT constraint such that the optimization or maximization solely determines the sign of the predictor in this case.

and HT hyperbola that is free of boundary points, see the appendix for details. Therefore, the solution $\mathbf{b}$ of the SSA problem must be a solution to the stationary Lagrangian equation

$$\boldsymbol{\gamma}_\delta = \tilde{\lambda}_1 2\mathbf{b} + \tilde{\lambda}_2(\mathbf{M}+\mathbf{M}')\mathbf{b} = \tilde{\lambda}_1 2\mathbf{b} + \tilde{\lambda}_2 2\mathbf{M}\mathbf{b}. \tag{7}$$

Since $\tilde{\lambda}_2 \neq 0$ (non-degenerate case) we obtain

$$D\boldsymbol{\gamma}_\delta \quad = \quad \mathbf{N}\mathbf{b}, \tag{8}$$

where $\mathbf{N} := (2\mathbf{M} - \nu\mathbf{I})$, $D = 1/\tilde{\lambda}_2 \neq 0$[3] and $\nu = -2\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}$. Furthermore, Equation (8) can be written in terms of a difference equation

$$b_{k+1} - \nu b_k + b_{k-1} \quad = \quad D\gamma_{k+\delta}\ ,\ 0 \leq k \leq L-1, \tag{9}$$

assuming $b_{-1} = b_L = 0$. The eigenvalues of $\mathbf{N}$ are $2\lambda_i - \nu$ with corresponding eigenvectors $\mathbf{v}_i$. If $\mathbf{b}(\nu)$ is the solution to the SSA problem, then $\nu/2$ cannot be an eigenvalue of $\mathbf{M}$ since otherwise $\mathbf{N}$ in Equation (8) would map one of the eigenvectors to zero which would contradict the last regularity assumption, noting that $D \neq 0$. Therefore, $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, ..., L\}$, $\mathbf{N}^{-1}$ exists and $\mathbf{N}^{-1} = \mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'$, where the diagonal matrix $\mathbf{D}_\nu^{-1}$ has entries $\frac{1}{2\lambda_i - \nu}$. Solving for $\mathbf{b}$ in Equation (8) gives

$$\begin{aligned} \mathbf{b} &= D\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta & (10)\\ &= D\mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'\mathbf{V}\mathbf{w} \\ &= D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i & (11) \end{aligned}$$

as claimed. For a proof of Assertion (2) we look at

$$\rho(\nu) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)'\mathbf{M}\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)}{\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)'\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)} = \frac{\sum_{i=1}^{L}\frac{\lambda_i w_i^2}{(2\lambda_i-\nu)^2}}{\sum_{i=1}^{L}\frac{w_i^2}{(2\lambda_i-\nu)^2}}. \tag{12}$$

We infer that $\lim_{\nu\to 2\lambda_i}\rho(\nu) = \lambda_i$, $i = 1, ..., L$. Since $\lambda_L = -\rho_{max}(L)$ and $\lambda_1 = \rho_{max}(L)$, we conclude that the limits $\pm\rho_{max}(L)$ can be reached by $\rho(\nu)$. Continuity of $\rho(\nu)$ and the intermediate-value theorem then imply that any $\rho_1 \in]-\rho_{max}(L), \rho_{max}(L)[$ is admissible for the HT constraint (the boundary cases where discussed in Corollary (1)).
We now proceed to a proof of Assertion (3):

$$\begin{aligned} \frac{d\rho\Big(y(\nu),y(\nu),1\Big)}{d\nu} &= \frac{d}{d\nu}\left(\frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}\right) = \frac{d}{d\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-1\ \prime}\mathbf{M}\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-1\ \prime}\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta}\right) = \frac{d}{d\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{M}\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta}\right) \\ &= \frac{2\boldsymbol{\gamma}_\delta'\mathbf{M}\mathbf{N}^{-3}\boldsymbol{\gamma}_\delta\cdot\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta - 2\boldsymbol{\gamma}_\delta'\mathbf{M}\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\cdot\boldsymbol{\gamma}_\delta'\mathbf{N}^{-3}\boldsymbol{\gamma}_\delta}{(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta)^2} & (13)\\ &= \frac{(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta)^2 - \boldsymbol{\gamma}_\delta'\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta\cdot\boldsymbol{\gamma}_\delta'\mathbf{N}^{-3}\boldsymbol{\gamma}_\delta}{(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta)^2}, & (14) \end{aligned}$$

where $\mathbf{N}^{-k} := (\mathbf{N}^{-1})^k$, $(\mathbf{N}^{-1})' = \mathbf{N}^{-1}$ (symmetry); also we relied on generic matrix differentiation rules in deriving Equation (13)[4]; finally, we inserted $2\mathbf{M}\mathbf{N}^{-k} = \mathbf{N}^{-k+1} + \nu\mathbf{N}^{-k}$ into the numerator of Equation (13) to obtain the last equation after simplification. Let now $\mathbf{N}^{-k} = \mathbf{V}\mathbf{D}^{-k}\mathbf{V}'$, where

[3] $\mathbf{b}$ is defined on a $L-2 \geq 1$-dimensional space so that the objective function is not overruled by the constraint, i.e., $|\tilde{\lambda}_2| < \infty$.

[4] $\frac{d(\mathbf{N}^{-1})}{d\nu} = \mathbf{N}^{-2}$ and $\frac{d(\mathbf{N}^{-2})}{d\nu} = 2\mathbf{N}^{-3}$. For the first equation the general rule is $\frac{d(\boldsymbol{\nu}^{-1})}{d\nu} = -\boldsymbol{\nu}^{-1}\frac{d\boldsymbol{\nu}}{d\nu}\boldsymbol{\nu}^{-1}$, noting that $\frac{d\boldsymbol{\nu}}{d\nu} = -\mathbf{I}$. The second equation follows by inserting the first equation into $\frac{d(\boldsymbol{\nu}^{-2})}{d\nu} = \frac{d(\boldsymbol{\nu}^{-1})}{d\nu}\mathbf{N}^{-1} + \mathbf{N}^{-1}\frac{d(\mathbf{N}^{-1})}{d\nu}$.

$\mathbf{D}^{-k}$, $k = 1, 2, 3$, is diagonal with eigenvalues $\lambda_{i\nu}^{-k} := (2\lambda_i - \nu)^{-k}$: the eigenvalues are (strictly) positive, if $\nu < -2\rho_{max}(L)$; if $\nu > 2\rho_{max}(L)$ then the eigenvalues are either (strictly) negative, if k is an odd number, or (strictly) positive, if k is an even number. For the numerator in Equation (14) we then obtain

$$\begin{aligned}
(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta)^2 - \boldsymbol{\gamma}_\delta'\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta \cdot \boldsymbol{\gamma}_\delta'\mathbf{N}^{-3}\boldsymbol{\gamma}_\delta &= (\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-2}\mathbf{V}'\boldsymbol{\gamma}_\delta)^2 - \boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-1}\mathbf{V}'\boldsymbol{\gamma}_\delta \cdot \boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-3}\mathbf{V}'\boldsymbol{\gamma}_\delta \\
&= (\mathbf{w}'\mathbf{D}^{-2}\mathbf{w})^2 - \mathbf{w}'\mathbf{D}^{-1}\mathbf{w} \cdot \mathbf{w}'\mathbf{D}^{-3}\mathbf{w} \\
&= \left(\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-2}\right)^2 - \sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-3}\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-1} \\
&= \sum_{i>k} w_i^2 w_k^2\left(2\lambda_{i\nu}^{-2}\lambda_{k\nu}^{-2} - \lambda_{i\nu}^{-1}\lambda_{k\nu}^{-3} - \lambda_{i\nu}^{-3}\lambda_{k\nu}^{-1}\right) \\
&= -\sum_{i>k} w_i^2 w_k^2\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-1}\left(\lambda_{i\nu}^{-1} - \lambda_{k\nu}^{-1}\right)^2 \;<\; 0,
\end{aligned}$$

where the strict inequality holds because $\lambda_{i\nu}^{-1} = (2\lambda_i - \nu)^{-1}$ are of the same sign, pairwise different and non-vanishing, if $|\nu| > 2\rho_{max}(L)$, and because $w_i \neq 0$ (completeness). Therefore, the numerator in Equation (14) is strictly negative and we conclude that $\rho\Big(y(\nu), y(\nu), 1\Big)$ must be a strictly monotonic function of ν for $\nu \in \{x|x > 2\rho_{max}(L)\}$ or for $\nu \in \{x|x < -2\rho_{max}(L)\}$. From $\lim_{|\nu|\to\infty}\rho(\nu) = \frac{\sum_{i=1}^L \lambda_i w_i^2}{\sum_{i=1}^L w_i^2} = \rho_{MSE}$ and $\frac{d\rho(\nu)}{d\nu} < 0$ we then infer $\max_{\nu<-2\rho_{max}(L)}\rho(\nu) = \min_{\nu>2\rho_{max}(L)}\rho(\nu) = \rho_{MSE}$, as claimed.
For a proof of the last Assertion (4) we look at

$$\begin{aligned}
\rho(y(\nu), z, \delta) &= \frac{\mathbf{b}'\boldsymbol{\gamma}_\delta}{\sqrt{\mathbf{b}'\mathbf{b}\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}} = D\frac{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta}{\sqrt{D^2\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}} = D\frac{\sum_{i=1}^L \frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i'\sum_{j=1}^L w_j\mathbf{v}_j}{\sqrt{D^2\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}} \\
&= \mathrm{sign}(D)\frac{\sum_{i=1}^L \frac{w_i^2}{2\lambda_i-\nu}}{\sqrt{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}.
\end{aligned}$$

For $\nu < -2\rho_{max}(L)$ the quotient is strictly positive and therefore positiveness of the objective function for the SSA solution implies $\mathrm{sign}(D) = 1$; for $\nu > 2\rho_{max}(L)$ the quotient is strictly negative so that $\mathrm{sign}(D) = -1$. Assume now that $\nu < -2\rho_{max}(L)$ so that

$$\begin{aligned}
\frac{d\rho\Big(y(\nu), z, \delta\Big)}{d\nu} &= \frac{d}{d\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta}{\sqrt{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\right) \\
&= \frac{1}{\left(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\right)^{3/2}\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\left\{\left(\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta\right)^2 - \boldsymbol{\gamma}_\delta'\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\mathbf{N}^{-3}\boldsymbol{\gamma}_\delta\right\} \\
&= \frac{\sqrt{\boldsymbol{\gamma}_\delta'\mathbf{N}^{-2}\boldsymbol{\gamma}_\delta}}{\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\frac{d\rho\Big(y(\nu), y(\nu), 1\Big)}{d\nu} < 0.
\end{aligned}$$

The last equality is obtained by recognizing that the expression in curly brackets is the same as the numerator of Equation (14). The proof applies also to $\nu > 2\rho_{max}(L)$ but with a changed sign, $\mathrm{sign}(D) = -1$, and accordingly modified strict inequality, as was to be shown. □

Remarks: The theorem derives exact finite-length solutions for any L such that $3 \leq L \leq T-1$ and the best predictor is obtained for $L = T - 1$ but the corresponding sample history would consist of a single point, thus rendering direct comparisons with a benchmark impossible. Also, Equations (4) and (9) correspond to frequency-domain and time-domain solutions whose peculiar structures will be explored and exploited further down. Finally, Equation (6) formalizes a trade-off or dilemma between the target correlation and the lag-one ACF for the solution of the SSA

problem which will be addressed and quantified in our examples.

The singular case of incomplete spectral support, assuming violation of the last regularity assumption of Theorem (1), is addressed in the following corollary.

Corollary 2. *Let all regularity assumptions of Theorem* (1) *hold, except completeness.*

1. *If* $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, ..., L\}$*, then the SSA predictor becomes*

$$\mathbf{b}(\nu) = D \sum_{i \in NZ} \frac{w_i}{2\lambda_i - \nu} \mathbf{v}_i, \tag{15}$$

where $NZ \subset \{1, ..., L\}$*. The lag-one ACF is*

$$\rho(\nu) = \frac{\sum_{i \in NZ} \frac{\lambda_i w_i^2}{(2\lambda_i - \nu)^2}}{\sum_{i \in NZ} \frac{w_i^2}{(2\lambda_i - \nu)^2}} =: \frac{M_1}{M_2}, \tag{16}$$

where M_1, M_2 *are identified with numerator and denominator in this expression.*

2. *Let* $\nu = \nu_{i_0} := 2\lambda_{i_0}$ *where* $i_0 \notin NZ$ *with adjoined rank-defficient* $\mathbf{N}_{i_0} = 2\mathbf{M} - \nu_{i_0}\mathbf{I}$*. Consider* $\mathbf{b}(\nu_{i_0})$*,* $\rho(\nu_{i_0})$ *and* $M_{i_0 1}, M_{i_0 2}$ *as defined in the previous assertion. In this case,* $\mathbf{b}(\nu_{i_0})$ *can be 'spectrally completed' as in*

$$\mathbf{b}_{i_0}(\tilde{N}_{i_0}) := \mathbf{b}(\nu_{i_0}) + D\tilde{N}_{i_0}\mathbf{v}_{i_0} \tag{17}$$

for some $\tilde{N}_{i_0}$ *to be determined below and with lag-one ACF*

$$\rho_{i_0}(\tilde{N}_{i_0}) = \frac{M_{i_0 1} + \lambda_{i_0}\tilde{N}_{i_0}^2}{M_{i_0 2} + \tilde{N}_{i_0}^2}. \tag{18}$$

If i_0 *is such that* $0 < \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} < \rho_1 < \lambda_{i_0}$ *or* $0 > \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} > \rho_1 > \lambda_{i_0}$*, then*

$$\tilde{N}_{i_0} \quad = \quad \pm\sqrt{\frac{\rho_1 M_{i_0 2} - M_{i_0 1}}{\lambda_{i_0} - \rho_1}} \tag{19}$$

ensures compliance with the HT constraint, i.e., $\rho_{i_0}(\tilde{N}_{i_0}) = \rho_1$*. The 'correct' sign-combination of* D *and* $\tilde{N}_{i_0}$ *is determined by the corresponding maximum of the SSA objective function.*

3. *If* $\boldsymbol{\gamma}_\delta$ *is not band limited, then any* ρ_1 *such that* $|\rho_1| \leq \rho_{max}(L)$ *is admissible in the HT constraint; If* $w_1 = 0, w_L \neq 0$ *then any* ρ_1 *such that* $-\rho_{max}(L) \leq \rho_1 < \rho_{max}(L)$ *is admissible ; If* $w_1 \neq 0, w_L = 0$ *then any* ρ_1 *such that* $-\rho_{max}(L) < \rho_1 \leq \rho_{max}(L)$ *is admissible; finally, if* $w_1 = w_L = 0$ *then any* ρ_1 *such that* $-\rho_{max}(L) < \rho_1 < \rho_{max}(L)$ *is admissible.*

A proof of the corollary and a worked-out example are provided in the appendix. The Corollary states that the domain of definition of ν can be extended to $\nu_{i_0} := 2\lambda_{i_0}$ (if $i_0 \notin NZ$) so that $\mathbf{N}_{i_0}$ becomes rank deficient (singular case). We can then augment and 'enrich' the ordinary solution $\mathbf{b}(\nu_{i_0})$ of Theorem (1) with the eigenvector $\mathbf{v}_{i_0}$ (in the Null-space of $\mathbf{N}_{i_0}$) in Equation (17) and we can select the weight $\tilde{N}_{i_0}$ to match the HT-constraint. In summary, the solution space increases and the spectrally completed $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ in Equation (17) can match HT constraints which are out of the range of the ordinary $\mathbf{b}(\nu_{i_0})$, previously obtained in the Theorem, see the Appendix for evidence. We now spell out explicitly the solution to the SSA problem based on the previous results.

Corollary 3. *Let the assumptions of theorem 1 hold. Then the solution to the SSA-optimization problem* (1) *is given by* $s\mathbf{b}(\nu_1)$ *where* $\mathbf{b}(\nu_1)$ *is obtained from Equation* (4)*, assuming an arbitrary scaling* $|D| = 1$ *(the sign of* D *is determined by asking for a positive objective function),* ν_1 *is a solution to the non-linear equation* $\rho(\nu_1) = \rho_1$ *and where* $s = \sqrt{l/\mathbf{b}(\nu_1)'\mathbf{b}(\nu_1)}$*. If the search for an optimal* ν_1 *can be restricted to* $\{\nu | |\nu| > 2\rho_{max}(L)\}$*, then* ν_1 *is determined uniquely by* ρ_1*.*

A proof follows directly from Theorem (1), noting that the scaling $s = \sqrt{l/\mathbf{b}(\nu_1)'\mathbf{b}(\nu_1)}$ interferes neither with the objective function nor with the HT constraint and can be set afterward, once a solution assuming an arbitrary scaling $|D| = 1$ has been obtained. In particular, Assertion (3) warrants uniqueness in $\{\nu||\nu| > 2\rho_{max}(L)\}$. □

A similar result can be obtained in the singular case addressed by Corollary (2). In addition to uniqueness, Assertion (3) also ensures swift numerical optimization due to strict monotonicity[5]. In this context, we now propose an exact closed-form solution for ν_1 as a function of ρ_1 in the non-linear HT equation $\rho(\nu_1) = \rho_1$ in the case of an AR(1) target specification.

Corollary 4. *Let the following assumptions hold in addition to the set of regularity conditions of theorem 1:*

1. $\boldsymbol{\gamma}_\delta \propto (\lambda^k)_{k=0,\ldots,L-1}$ *with stable root* $\lambda \neq 0$ *(stationary AR(1));*
2. $|\nu| > 2$ *so that* $\nu = \lambda_{\rho_1} + 1/\lambda_{\rho_1}$ *where* $\lambda_{\rho_1} \in]-1, 1[\backslash\{0\}$*;*
3. $\lambda \neq \lambda_{\rho_1}$ *(non-singular case);*
4. $\lambda_{\rho_1}, \lambda$ *and* L *are such that* $\max(|\lambda_{\rho_1}|^{2k}, |\lambda|^{2k})$ *is negligible for* $k > L$ *(sufficiently fast decay for given* L*).*

Then the optimal λ_{ρ_1} *is obtained as (the real-valued)*

$$\lambda_{\rho_1} = -\frac{1}{3c_3}\left(c_2 + C + \frac{\Delta_0}{C}\right), \tag{20}$$

where

$$\begin{aligned} C &= \sqrt[3]{\frac{\Delta_1 + \text{sign}(\Delta_1)\sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\ \Delta_0 &= c_2^2 - 3c_3c_1 \\ \Delta_1 &= 2c_2^3 - 9c_3c_2c_1 + 27c_3^2c_0 \end{aligned} \tag{21}$$

and where c_3, c_2, c_1, c_0 *are the coefficients of a cubic polynomial which depend on the AR(1)-target specified by* λ *and the HT constraint* ρ_1 *according to*

$$c_3 = \lambda^{-2} - \rho_1\lambda^{-1}\ ,\ c_2 = -\lambda^{-1} - \rho_1(\lambda^{-2} - 2)\ ,\ c_1 = -1 - \rho_1(\lambda - 2\lambda^{-1})\ ,\ c_0 = \lambda - \rho_1.$$

The SSA predictor $\mathbf{b}(\nu_1) = \mathbf{b}(\lambda_{\rho_1} + 1/\lambda_{\rho_1})$ *is then uniquely determined in closed-form by 4, down to a proper scaling, ensuring compliance with the length constraint, and the correct sign, leading to a positive criterion value* $\mathbf{b}(\nu_1)'\boldsymbol{\gamma}_\delta > 0$*.*

A proof of the corollary exploits the (time-domain) difference Equation (9) and can be found in the appendix. Note that a closed-form expression in the case of an AR(p) target with $p > 1$ does generally not exist, see the appendix for reference. We now address the distribution of the SSA predictor.

Corollary 5. *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* $\boldsymbol{\mu}_{\gamma_\delta}$ *and variance* $\boldsymbol{\Sigma}_{\gamma_\delta}$*. Then mean and variance of the SSA predictor* $\hat{\mathbf{b}}(\nu)$ *are*

$$\begin{aligned} \boldsymbol{\mu}_{\hat{\mathbf{b}}} &= D\mathbf{N}^{-1}\boldsymbol{\mu}_{\gamma_\delta} \\ \boldsymbol{\Sigma}_{\hat{\mathbf{b}}} &= D^2\mathbf{N}^{-1}\boldsymbol{\Sigma}_{\gamma_\delta}\mathbf{N}^{-1}. \end{aligned}$$

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

[5]The optimal parameter can be obtained by triangulation, in intervals of exponentially decaying width of order $1/2^n$, where n is the number of iteration steps.

The proof follows directly from Equation 4 and we refer to standard textbooks for a derivation of mean, variance and distribution of the MSE-estimate under various assumptions about x_t, see Brockwell and Davis (1993). Our last result in this section proposes a dual interpretation of the SSA predictor.

Corollary 6. *Let all regularity assumptions of Theorem* (1) *hold and let* $y_t(\nu_1)$ *denote the SSA-solution for some* $\nu_1 > 2\rho_{max}(L)$. *Set* $\rho_{\nu_1,\delta} := \rho(y(\nu_1), z, \delta)$ *and consider the dual optimization problem*

$$\left.\begin{array}{c} \max_{\mathbf{b}} \rho(y, y, 1) \\ \rho(y, z, \delta) = \rho_{\nu_1,\delta} \end{array}\right\}. \tag{22}$$

The solution to this criterion has the same functional form as the original SSA problem and if the search for an optimal ν *can be restricted to the set* $\{\nu||\nu| > 2\rho_{max}(L)\}$ *then* $y_t(\nu_1)$ *is also the solution to the dual problem. If* $\nu_1 < -2\rho_{max}(L)$, *then* $y_t(\nu_1)$ *is the solution to the dual problem if minimization is substituted for maximization in the objective of Criterion* (22).

Proof: Let us first examine the case $\nu_1 > 2\rho_{max}(l)$. The Lagrangian Equation (7) does not discern constraint and objective: after suitable re-scaling of multipliers, the problem specified by Criterion (22) leads to the same functional form $\mathbf{b} = D\mathbf{N}^{-1}\boldsymbol{\gamma}_\delta$ of its solution[6]. The only difference is that ν in Criterion (22) must be selected such that $\rho(y(\nu), z, \delta) = \rho_{\nu_1,\delta}$. If the search can be restricted to $\{\nu|\nu > 2\rho_{max}(L)\}$, then by Assertion (4) the solution to the primal problem is also a solution to the dual problem, due to strict monotonicity of $\rho(y(\nu), z, \delta)$. The extension to $\{\nu||\nu| > 2\rho_{max}(L)\}$ follows from Assertion (3) which affirms that $\rho(\nu) < \rho(\nu_1)$ if $\nu < -2\rho_{max}(L)$. A similar reasoning applies if $\nu_1 < -2\rho_{max}(L)$, noting that maximization must be replaced by minimization in the dual Criterion (22) (because $\rho(\nu) > \rho(\nu_1)$ if $\nu > 2\rho_{max}(L)$). □

This last result distinguishes the SSA solution as the predictor with the largest HT (lag-one ACF) for a given target correlation or tracking accuracy. We now interpret the SSA solution and we link HT and (lag-one) ACF. In particular, we shall see that the restriction $|\nu| > 2\rho_{max}(L)$ in the previous Corollaries (3), (4) and (6) is not a limitation in typical applications.

4 Interpretation

4.1 Zero-Crossings, Sign-Accuracy and a Link between HT and (Lag-One) ACF

Assume ϵ_t to be Gaussian noise and let $\mathrm{SA}(y_t) := \mathrm{P}\Big(\operatorname{sign}(z_{t+\delta}) = \operatorname{sign}(y_t)\Big)$ denote the probability of same sign of target and predictor, where the acronym SA refers to Sign Accuracy. Gaussianity then implies

$$SA(y_t) = 2E\left[I_{\{z_{t+\delta}\geq 0\}} I_{\{y_t \geq 0\}}\right] = 0.5 + \frac{\arcsin(\rho(y, z, \delta))}{\pi} \tag{23}$$

so that maximization of the target correlation $\rho(y, z, \delta)$ or of SA are equivalent optimization principles. The link to SA motivated the choice of the objective function of Criterion (2) in the first place, see Wildi (2024). Proceeding further, we now define HT 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., $t_{i-1} < t_i$, $t_1 \geq L$, $\operatorname{sign}(y_{t_i-1} y_{t_i}) < 0$ for all i and $\operatorname{sign}(y_{t-1} y_t) > 0$ if $t_{i-1} < t < t_i$. Under the above stationarity assumptions, $ht(y|\mathbf{b}, i) = ht(y|\mathbf{b})$ measures the expected duration between consecutive zero-crossings of y_t, see Kedem (1986).

Proposition 2. *Let* y_t *be a zero-mean stationary Gaussian process. Then the HT* $ht(y|\mathbf{b})$ *is*

$$ht(y|\mathbf{b}) = \frac{\pi}{\arccos(\rho(y, y, 1))}. \tag{24}$$

[6]Re-scaling is always possible because the regularity assumptions imply non-vanishing and finitely-sized multipliers.

We refer to Kedem (1986) for proof. The bijective link between the HT and the lag-one autocorrelation in Equation (24) suggests that Criterion (1) can be interpreted as a maximization of SA under a fixed expected rate of zero crossings of the predictor

$$\begin{aligned} &\max_{\mathbf{b}} SA(y_t) \\ &ht(y|\mathbf{b}) = ht_1. \end{aligned}$$

Interpreted in its dual form, the predictor generates the fewest crossings for a given sign-accuracy, see Corollary (6).

Since zero-crossings, sign accuracy and correlations are indifferent to the scaling of the filter, we can select $s_{MSE} := \mathbf{b}'\boldsymbol{\gamma}_\delta / \mathbf{b}'\mathbf{b}$ such that MSE performances of $s_{MSE}\mathbf{b}$ are optimized conditional on the imposed HT constraint. In this context, we could look at the alternative SSA-MSE criterion

$$\left.\begin{array}{c} \min_{\mathbf{b}} (\boldsymbol{\gamma}_\delta - \mathbf{b})'(\boldsymbol{\gamma}_\delta - \mathbf{b}) \\ \mathbf{b}'\mathbf{M}\mathbf{b} = \mathbf{b}'\mathbf{b}\rho_1 \end{array}\right\}, \tag{25}$$

where the objective function $(\boldsymbol{\gamma}_\delta - \mathbf{b})'(\boldsymbol{\gamma}_\delta - \mathbf{b})$ is the MSE and the length constraint is dropped. The corresponding Lagrangian heads to a system of equations $2(\boldsymbol{\gamma}_\delta - \mathbf{b}) = 2\tilde{\lambda}(\mathbf{M} - \rho_1\mathbf{I})\mathbf{b}$ or, more compactly

$$\boldsymbol{\Psi}^{-1}\mathbf{b} = F\boldsymbol{\gamma}_\delta,$$

where $\boldsymbol{\Psi} = (2\mathbf{M} - \psi\mathbf{I})$ with $\psi = 2(\rho_1 - 1/\tilde{\lambda})$, $F = 2/\tilde{\lambda}$ and where ψ can be selected for compliance with the HT constraint. In summary, SSA reconciles MSE, sign accuracy, and smoothing requirements in a flexible and interpretable way.

To conclude, we note that target and predictor can be nearly Gaussian, due to aggregation by the filter (central limit theorem), even if ϵ_t isn't, so that the above transformations, linking correlations, HT and SA, might still be practically relevant despite the violation of the Gaussian assumption: an example is proposed in the appendix and Wildi (2024) shows the resilience of the HT Equation (24) for an application to the S&P-500 Index, whose log-returns conflict overtly with the Gaussian assumption.

4.2 Frequency-Domain Analysis and an Application To Real-Time Signal Extraction

Formally, the SSA-AR(2) filter in the difference Equation (9) has transfer function $\Gamma_{AR(2)}(\nu, \omega) = \frac{1}{\exp(-i\omega) - \nu + \exp(i\omega)} = \frac{1}{2\cos(\omega) - \nu}$. Let then $\boldsymbol{\Gamma}_{AR(2)}(\nu)$ denote the vector of transfer function ordinates of SSA-AR(2) evaluated at the Fourier frequencies $\omega_j = j\pi/(L+1)$, $j = 1, ..., L$. Equation (11) implies

$$\mathbf{b}(\nu)'\boldsymbol{\epsilon}_t = D(\nu, l) \sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu} \mathbf{v}_i'\boldsymbol{\epsilon}_t, \tag{26}$$

with $\lambda_i = \cos(\omega_i)$ and where $\mathbf{v}_i'\boldsymbol{\epsilon}_t$ is the projection of the data on the i-th Fourier vector $\mathbf{v}_i$. The weights assigned to these projections by $\mathbf{b}(\nu)$ are $\mathbf{w} \odot \boldsymbol{\Gamma}_{AR(2)}(\nu)$ where $\odot$ designates the Hadamard product: this expression corresponds to the convolution of SSA-AR(2) and $\boldsymbol{\gamma}_\delta$ in the frequency domain. We then refer to $|\mathbf{w}|$ and $\left|\mathbf{w} \odot \boldsymbol{\Gamma}_{AR(2)}(\nu)\right| = |\mathbf{w}| \odot \left|\boldsymbol{\Gamma}_{AR(2)}(\nu)\right|$ in terms of (SSA-) amplitude functions of $\boldsymbol{\gamma}_\delta$ and $\mathbf{b}(\nu)$, respectively. Moreover

$$(\mathbf{V}'\mathbf{b}(\nu))'\boldsymbol{\epsilon}_t = D(\nu, l) \sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu} \mathbf{e}_i'\boldsymbol{\epsilon}_t = D(\nu, l) \sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu} \epsilon_{t+1-i},$$

where $\mathbf{e}_i$ is the i-th unit vector, can be interpreted as a discrete Fourier transform (SSA-DFT) and its square as an SSA periodogram of the predictor. Also, from Equation (4) $\mathbf{b}'(\nu)\mathbf{b}(\nu) =$

$D(\nu,l)^2\sum_{i=1}^{L}\left(\frac{w_i}{2\lambda_i-\nu}\right)^2$ is Parseval's identity and $D(\nu,l)^2\left(\frac{w_i}{2\lambda_i-\nu}\right)^2$ measures the contribution of $\mathbf{v}_i$ to the variance of the predictor. These results may be used to link SSA to the Direct Filter Approach (DFA) proposed by McElroy and Wildi (2016). For $\nu \leq -2$, the function $|2\cos(\omega)-\nu|$ is monotonically decreasing for $\omega \in [0,\pi]$ and therefore SSA-AR(2) is a highpass with a peak of its transfer (or amplitude) function at frequency π (if $\nu=-2$ then the peak is infinite); if $\nu \to -\infty$ then $D(\nu,l)\mathbf{\Gamma}_{AR(2)}(\nu)$ becomes asymptotically an allpass and $\mathbf{b}(\nu) \to \sqrt{l}\boldsymbol{\gamma}_\delta/\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, i.e., the SSA predictor converges to the scaled MSE predictor of variance l (degenerate case excluded by Theorem (1)). The highpass favors high-frequency (noise) leakage and increases the number of zero-crossings, as requested when $ht_1 < ht_{MSE}$, or $\rho_1 < \rho_{MSE}$, in the HT constraint. For $\nu \geq 2$, the function $|2\cos(\omega)-\nu|$ is monotonically increasing for $\omega \in [0,\pi]$ and accordingly SSA-AR(2) is a lowpass with a peak of its transfer or amplitude function at frequency zero: the filter damps high-frequency noise and lessens the number of zero-crossings, as requested when $ht_1 > ht_{MSE}$ in the HT constraint. For $-2 < \nu < 2$, SSA-AR(2) is a bandpass with a peak of its transfer or amplitude function at frequencies $\omega = \pm\arccos(\nu/2)$.

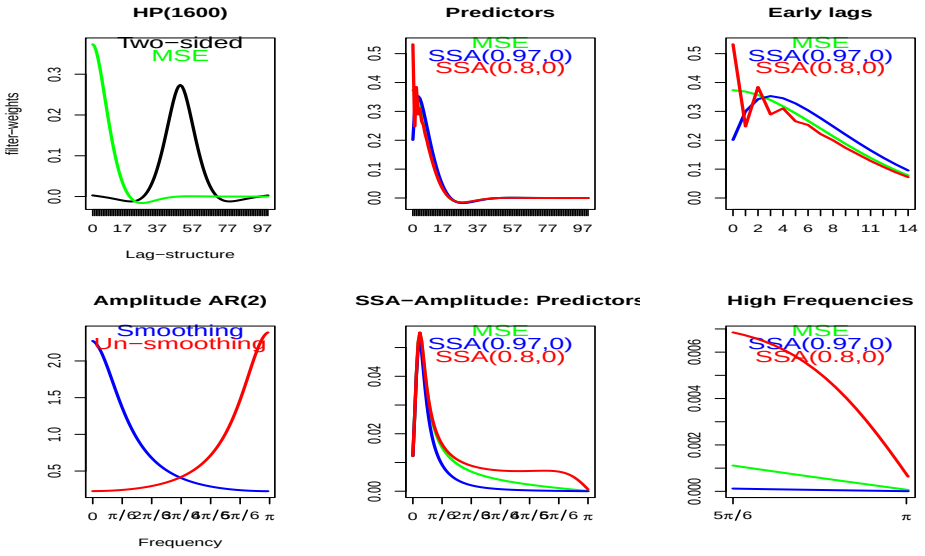

Figure 1: Two-sided truncated HP(1600), centered at lag $k = 50$ (black), and three nowcasts: MSE (green), SSA(0.97,0) (blue) and SSA(0.8,0) (red). All filters are arbitrarily scaled to unit length (unit variance when fed with standardized WN). Filter coefficients (top graphs) and SSA-amplitude functions (bottom graphs). The first few lags are highlighted in the top rightmost plot. Amplitude of SSA-AR(2) (bottom left), of nowcasts (bottom center) and high frequencies (bottom right). All SSA-amplitude functions are artificially aligned at frequency zero.

For illustration, we apply the SSA criterion to the quarterly HP filter with parameter $\lambda = 1600$, see Hodrick and Prescott (1997), with two-sided (bi-infinite symmetric) target γ_k displayed in Figure (1): the two-sided filter (black line) has been truncated and right-shifted (centered at lag $k = 50$) for ease of visual inspection. The HP trend filter can be interpreted as an optimal MSE-signal extraction filter for the trend in the smooth trend model, see Harvey (1989). We aim at approximating the HP target $z_{t+\delta}$ for $\delta = 0$ by a nowcast y_t based on a one-sided filter b_k, $k = 0,...,100$, of length $L = 101$. The MSE nowcast $\boldsymbol{\gamma}_0$ corresponds to the right tail of the two-sided filter and has lag-one ACF $\rho_{MSE} = \boldsymbol{\gamma}_0'\mathbf{M}\boldsymbol{\gamma}_0/\boldsymbol{\gamma}_0'\boldsymbol{\gamma}_0 = 0.926$. We compute two SSA nowcasts, imposing lag-one ACFs of $0.97 > \rho_{MSE}$ (smoothing) and $0.8 < \rho_{MSE}$ (un-smoothing) with resulting $\nu_1 = 2.44 > 2$ and $\nu_2 = -2.42 < -2$, see Fig.(1). Optimal smoothing and un-smoothing

are obtained by lowpass ($\nu_1 > 2$) and highpass ($\nu_2 < -2$) AR(2)-filters, respectively, and the SSA-amplitude functions $\left|D(\nu_i, l)\mathbf{\Gamma}_{AR(2)}(\nu_i) \odot \mathbf{w}\right|$, $i = 1, 2$, of the corresponding SSA designs are below or above the amplitude $|\mathbf{w}|$ of the MSE benchmark towards higher frequencies, assuming an artificial alignment of all amplitude functions at frequency zero for better visual inspection.

The SSA-amplitude functions in the bottom-mid panel of the above figure suggests evidence of a strong peak to the right of frequency zero. As we shall see, this peak is an artifact due to the particular frequency-domain (FD) basis $\mathbf{V}$ derived from the eigenvectors of $\mathbf{M}$. To begin, we note that $\mathbf{v}_j$, $j = 1, ..., L$, correspond to the imaginary part of the 'classic' complex-valued FD-basis $\exp(ij\boldsymbol{\omega})$, where $\boldsymbol{\omega} = (\pi/(L+1), ..., L\pi/(L+1))'$. In contrast to the latter, the SSA-basis $\mathbf{v}_j$ complies with the boundary constraints imposed upon $\mathbf{b}(\nu)$ since $\sin(kj\pi/(L+1)) = b_{k-1}(\nu) = 0$ for $k \in \{0, L+1\}$. We can now compute and compare 'classic' and SSA-amplitude functions, see Fig. (2).

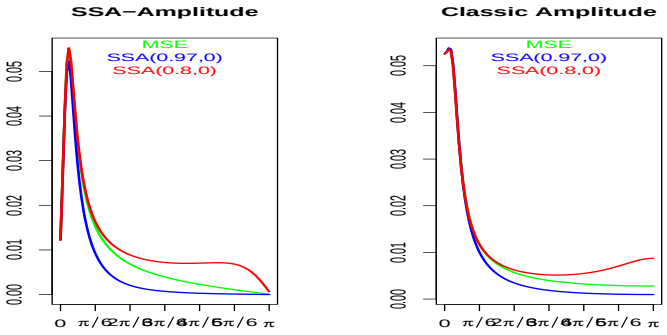

Figure 2: Comparison of SSA-amplitude (left) and classic amplitude functions (right).

The differences in the figure are reliant solely on the choice of the orthonormal basis for the frequency domain decomposition, $\mathbf{v}_j$ to the left and $\exp(ij\boldsymbol{\omega})$ to the right, and this choice affects the graphical interface for understanding, explaining or interpreting the solution to Criterion (1). In particular, the convolution result $\mathbf{w} \odot \mathbf{\Gamma}_{AR(2)}(\nu)$ that helps explain the action of the filter does not hold when exchanging bases, replacing $\mathbf{v}_j$ by $\exp(ij\boldsymbol{\omega})$. Clearly, the peaks of the SSA-amplitude functions to the right of frequency zero in the left panel are an artifact due to the boundary constraint $\sin(kj\pi/(L+1)) = 0$ at $k = 0$ (frequency zero) and a similar effect can be observed towards frequency π. To resolve the ambiguity, we propose to rely on $\mathbf{v}_j$ for exact spectral decomposition results, including convolution, DFT and Parseval equality, while an analysis of classic filter characteristics, such as amplitude or phase shift of the predictor, are better accommodated by swapping $\exp(ij\boldsymbol{\omega})$ for $\mathbf{v}_j$, as done in the right panel of Fig.(2).

To conclude we now briefly discuss the implications of the restriction $|\nu| > 2\rho_{max}(L)$ in the previous Corollaries. As shown, Equation (4) is the convolution of the SSA-AR(2) and the target, as decomposed in the SSA-basis $\mathbf{V}$. If $|\nu| \leq 2$, then the function $|2\cos(\omega) - \nu|$ vanishes at $\omega_0 := \arccos(\nu/2)$ and the SSA-AR(2) in Equation (9) is a non-stationary filter with unit-roots at frequencies $\pm\omega_0$: we therefore refer to $\nu \in [-2, 2]$ as the *unit-root case*. For $|\nu| < 2$ the integration order of SSA-AR(2) is one, but for $|\nu| = 2$ the previously distinct roots merge such that the integration order becomes two. For large L, the eigenvalues $\lambda_j = \cos(\omega_j)$, $j = 1, ..., L$ of $\mathbf{M}$ are increasingly densely packed in $[-1, 1]$ and therefore $\lim_{L\to\infty} \max |\mathbf{\Gamma}_{AR(2)}(\nu)| = \infty$, i.e., the peak of the amplitude of $\mathbf{\Gamma}_{AR(2)}(\nu)$ diverges as the sample size increases, irrespective of $\nu \in [-2, 2]$. If the target has complete spectral support, we then deduce that the convolution $\mathbf{w} \odot \mathbf{\Gamma}_{AR(2)}(\nu)$ will generally lead to a corresponding spectral peak of $|\mathbf{V}'\mathbf{b}|$ and $\mathbf{b}$ becomes increasingly periodic for large L, see Fig.(3).

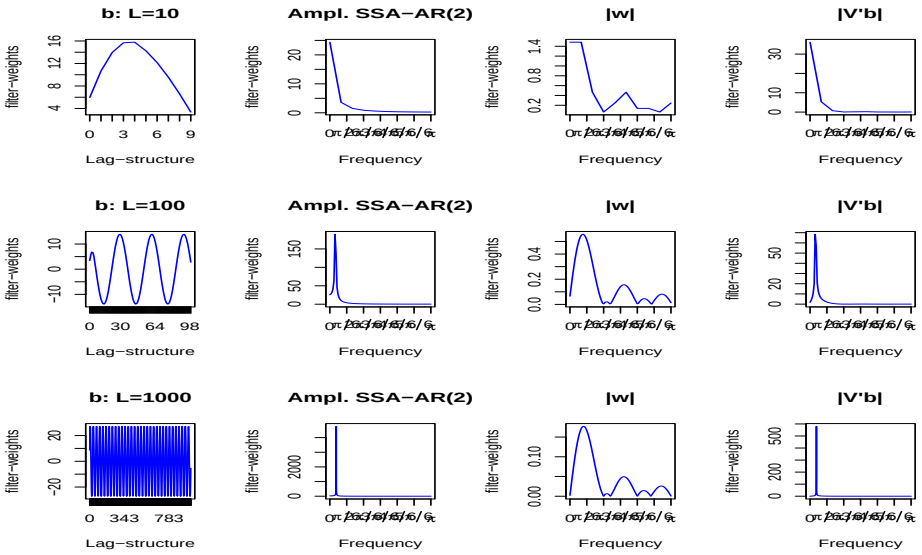

Figure 3: Coefficients of SSA-filters based on the HP trend target and a fixed $\nu = 1.96$ (leftmost panels) with corresponding SSA-AR(2) amplitude functions (second from left), SSA-amplitudes of target (third from left) and SSA-amplitudes of $\mathbf{b}(\nu)$ for lengths L=10 (top), L=100 (middle) and L=1000 (bottom). Amplitudes in the rightmost panels are identical to the product of the two amplitudes to the left, by convolution.

We display $\mathbf{b}(\nu)$ (leftmost), $|\mathbf{\Gamma}_{AR(2)}(\nu)|$ (second from left), $|\mathbf{w}|$ (third from left) and $|\mathbf{V}'\mathbf{b}|$ (rightmost) based on the above HP target for a fixed $\nu = 1.96 \in [-2, 2]$ and filter lengths $L = 10$ (top), 100 (middle) and 1000 (bottom). Note that $|\mathbf{V}'\mathbf{b}| = |\mathbf{w}| \odot |\mathbf{\Gamma}_{AR(2)}(\nu)|$, by convolution. For small filter lengths, $\mathbf{b}(\nu)$ may eventually comply with a (trend-) lowpass design but as L increases, the peak of SSA-AR(2) narrows and $\mathbf{b}(\nu)$ becomes increasingly periodic, with periodicity determined by the unit-root frequency $\omega_0 = \arccos(\nu/2) \approx \pi/15.68$. The cyclical pattern is due to convolution $\mathbf{w} \odot \mathbf{\Gamma}_{AR(2)}(\nu)$, noting that $|\mathbf{w}|$ does not vanish because the target has complete spectral support. In any case, the strongly periodic SSA-nowcast seems hardly reconcilable with the (lowpass) HP target specification for (any) $\nu \in [-2, 2]$ and L sufficiently large which suggests that the restriction $|\nu| > 2\rho_{max}(L)$ imposed by Corollaries (3), (4) and (6) could be replaced by the more stringent $|\nu| > 2$ and is not a limitation in typical applications. Note also that the case $|\nu| > 2$ refers to an unstable SSA-AR(2) whose characteristic polynomial has roots $\lambda, 1/\lambda$ such that $\nu = \lambda + 1/\lambda$. Interestingly, potential instability of Equation (9) is effectively contained by imposing the boundary constraints $b_{-1}(\nu) = b_L(\nu) = 0$.

4.3 Benchmark Customization

Table (1) compares target correlations $\rho(y, z, \delta)$, sign accuracies based on Equation (23), lag-one ACF $\rho(y, y, 1)$ and HTs based on Equation (24) of the filters in the previous section. A

	HP(1600)	MSE	SSA(0.97,0)	SSA(0.8,0)
Target correlation	1.000	0.733	0.717	0.716
SA	1.000	0.762	0.754	0.754
Lag one ACF	0.996	0.926	0.970	0.800
HT	34.316	8.138	12.793	4.882

Table 1: Target correlation, sign accuracy, lag-one ACF and HT of SSA designs applied to HP

comparison of the HTs of the target and MSE predictor in the first two columns suggests that the latter is subject to substantial leakage. Indeed, unwanted 'noisy' crossings of the predictor are often clustered in the vicinity of target crossings, when both filters hover over the zero line. We then argue that an explicit control of noisy crossings due to an unduly small HT of the (classic MSE) predictor, is a relevant objective, see Wildi (2024) for an application to real-time business cycle analysis. Moreover, Criterion (1) ensures an optimal tracking of the target by SSA: this property warrants that the interpretation or the economic content supported by z_t, such as, e.g., a business cycle indicator, can be transferred to SSA. Finally, SSA effectively minimizes, if $\nu > 2$, or maximizes, if $\nu < -2$, the rate of zero-crossings in the class of all predictors with the same target correlation (dual interpretation) and therefore Criterion (1) addresses the problem of noisy false alarms in some way optimally. If we now consider the MSE predictor in the above example as a benchmark for the particular HP nowcast problem, then we can interpret the application of SSA to HP in terms of a customized benchmark predictor with enhanced noise suppression if $\rho_1 > \rho_{MSE}$ in the HT constraint. This feature is exploited in Wildi (2024) who shows that the customized designs generate less 'false alarms' at the transitions from recession and expansion episodes while preceding the benchmark at the effective transition points. In addition to the HP-design, the SSA-package provides customizations of the Hamilton (2018) regression filter and of the Baxter-King (1999) filter.

5 Forecasting, Signal Extraction and a Prediction Trilemma

5.1 Forecasting

For illustration of the relevant facets of the SSA predictor, we are interested in deriving a one-step-ahead forecast for the MA(2)-process $z_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$, where $\gamma_k = 1, k = 0, 1, 2$, $\delta = 1$ and where ϵ_t are assumed to be known, mainly for simplicity of exposition. For this purpose, 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 HTs ht =3.74 and 10; the third filter deviates from the second one by selecting $L = 50$; the HT 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 (24); the second HT $ht = 10$ is sufficiently different in size to reveal some of the salient features of the approach. For our comparison we also include the MSE forecast $\hat{z}^{MSE}_{t+1} = \epsilon_t + \epsilon_{t-1}$ as well as a 'lag-by-one' or no-change forecast $\hat{z}^{lag\ 1}_{t+1} = z_t$, see Fig. (4) (an arbitrary scaling scheme is applied to SSA filters).

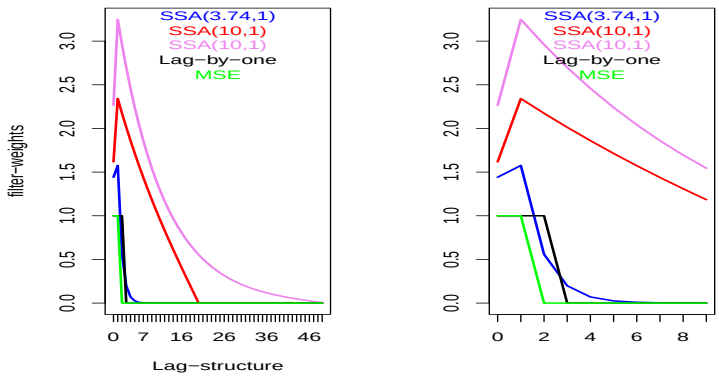

Figure 4: Coefficients of MSE-, SSA- and lag-by-one forecast filters with arbitrarily scaled SSA designs. All lags (left panel) and first ten lags (right panel).

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 HT constraint (stronger smoothing). For a fixed filter length L, a larger HT ht asks for a slower zero-decay of filter coefficients (blue vs. red lines) and for fixed HT ht, a larger L leads to a faster zero-decay but a longer 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 implicit boundary constraint $b_{-1} = 0$, see Theorem (1). Note that the 'lag-by-one' forecast (black) has the same HT as the first SSA filter (blue) so that the latter should outperform the former with respect to sign accuracy or, equivalently, in terms of target correlation with the shifted target, as confirmed in Table (2). MSE outperforms all other forecasts in terms of correlation and sign

	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
HT	3.735	10.000	10.000	3.735	3.000
Sign accuracy	0.788	0.626	0.627	0.732	0.804

Table 2: Performances of MSE and lag-by-one benchmarks vs. SSA. The two columns referring to SSA(10,1) correspond to filter lengths 20 (first) and 50 (second).

accuracy but it loses with respect to smoothness or HT; SSA(3.74,0) outperforms the lag-by-one benchmark; both SSA(10,0) loose in terms of sign accuracy but win with respect to smoothness and while the profiles of longer and shorter filters differ in Figure (4), their respective performances are virtually indistinguishable in Table (2), suggesting that the selection of L is to some extent uncritical, assuming it is at least twice the HT $L \geq 2ht_1$. The table also illustrates the tradeoff between target correlation (or sign accuracy) and HT, which is formalized by Equation (6). Table (3) allows for a more detailed analysis based on a finer grid of HT values. The SSA framework

	ht=4	ht=4.5	ht=5	ht=5.5	ht=6	ht=7	ht=8	ht=9	ht=10
Target 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 3: Tradeoff: effect of the HT on target correlation (first row) and sign accuracy (last row) for fixed forecast horizon.

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 tracking the level with higher accuracy (larger $\rho(y, z.\delta)$), but reduced smoothness (smaller HT), or for assessing sign changes more reliably. In the latter case, Corollary (6) asserts optimality in the sense that SSA effectively minimizes the rate of zero-crossings or alarms for a given tracking accuracy. Formally, a decision for one or several interesting designs could be based on re-computing the above 'tradeoff table' for the particular prediction problem at hand based on the provided SSA package.
In a final step, we allow the previously fixed forecast horizon $\delta = 1$ to vary, see Fig.(5) for illustration.

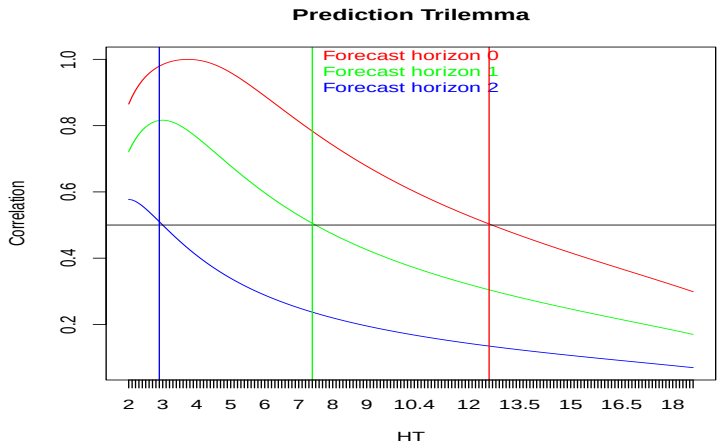

Figure 5: Target correlations of the SSA predictor as a function of the forecast horizon and the HT.

For each $\delta = 0, 1, 2$, the figure displays the target correlation for given HTs on the abscissa. The peak of the correlation for a given δ appears at a particular HT value which corresponds to the classic MSE predictor for that forecast horizon: for $\delta = 0$, $y_{t,MSE} = z_t$ and the peak target correlation, obtained at the HT 3.74 of z_t, is one. To the right of the peak, corresponding to values $\nu > 2$ in Theorem (1), SSA generates fewer crossings (stronger smoothing than MSE) and to its left more ($\nu < -2$); on both sides, SSA maximizes the target correlation, subject to the imposed HT constraint ht_1 on the abscissa; in its equivalent dual form, SSA maximizes the HT for given target correlation on the right of the peak ($\nu > 2$), see Corollary (6); finally, the correlation curve for $\delta = 1$ replicates entries in Table (3). 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}, δ_i)-predictors y_{ti} have a fixed correlation $\rho(y_i, z, \delta_i) = 0.5$ with the target $z_{t+\delta_i}$, marked by the horizontal black line which intersects the curves at the corresponding HTs ht_{1i}, $i = 1, 2, 3$. Fig.(5) generalizes the dilemma in Table (3) 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, δ) and classic MSE performances can be replicated by selecting $\delta = \delta_0$ and $ht_1 = ht_{MSE}$, the HT 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. These results together with the orthogonal frequency-domain decomposition in Section (4.2) may be used to establish a link to McElroy and Wildi (2019) who propose a trilemma enunciated in the frequency-domain, but in contrast to SSA, their smoothness component does not address zero-crossings explicitly which is a weakness in certain applications such as, e.g., BCA.

5.2 Real-Time Signal Extraction: Addressing Timeliness and Smoothness

We here rely on the HP(1600) target in Section (4.2) and compare performances of three SSA designs when nowcasting the acausal filter: SSA(0.926,0), SSA$(0.97, 0)$ and SSA$(0.97, 12)$. The first design, based on hyperparameters $(\rho_1 = \rho_{MSE}, \delta = 0)$ imposes the lag-one ACF of the MSE nowcast and thus replicates the latter; the second design, based on $(\rho_1 = 0.97, \delta = 0)$ emphasizes smoothness by imposing a larger lag-one ACF; finally, the third SSA based on $(\rho_1 = 0.97, \delta = 12)$ highlights smoothness as well as timeliness (relative lead). As discussed in the previous section, we

interpret all three designs as nowcasts of the two-sided HP(1600) emphasizing different priorities, i.e., our effective forecast horizon is $\delta_0 = 0$ and all target correlations are evaluated accordingly. Table (4) summarizes and compares expected and empirical performances, whereby the latter are based on a long sample of (Gaussian) white noise, and Fig.(6) compares a subsample of the cor-

	MSE	SSA(0.97,0)	SSA(0.97,12)
ht	8.138	12.793	12.793
Empirical HT	8.151	12.716	12.793
Target correlation	0.733	0.717	0.512
Emp. correlation	0.732	0.716	0.510

Table 4: HTs and target correlations of three SSA nowcasts based on different hyperparameter specifications: expected vs. empirical numbers, based on a sample of Gaussian noise of length 100000. All target correlations are computed at $\delta_0 = 0$ (nowcast).

responding filter outputs. Our results confirm the prediction trilemma enounced in the previous section: the first (MSE-) nowcast outperforms in terms of target correlation; the other two SSA designs emphasize smoothness equally and together outperform the MSE benchmark in terms of HT; the third nowcast outperforms the other two with respect to timeliness or relative advancement (red line in the figure) but it ranks last according to the target correlation. Interestingly, the timing of its zero-crossings often accords with the acausal target, in contrast to the other two nowcasts which are systematically delayed. In summary, our results confirm that the hyperparameter pairing (δ, ht) of SSA can accommodate various practically relevant research priorities in terms of smoothness, timeliness and target correlation or MSE.

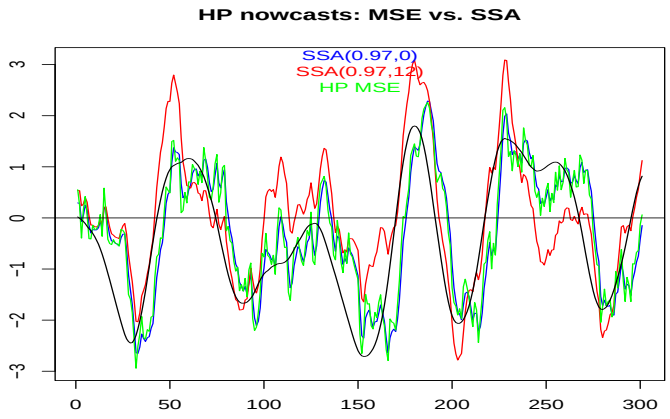

Figure 6: Acausal HP(1600) target (black) and nowcasts: MSE (green), SSA(0.97,0) (blue) and SSA(0.97,12) (red). Both SSA cross the zero line less frequently than MSE and the second SSA (red) is systematically left shifted with a mean lead time or advancement of roughly 1.5 time units when compared to the benchmark. Its zero crossings track the main up- and downturns of the acausal target in a more timely fashion than the other two nowcasts which are subject to a systematic retardation.

6 Autocorrelation

We here propose an extension of the simple WN framework to autocorrelated processes, distinguishing stationary and non-stationary integrated processes.

6.1 Stationary Processes

Consider the generalized target $\tilde{z}_t = \sum_{|k|<\infty} \gamma_k x_{t-k}$ where we assume $x_t = \sum_{i=0}^{\infty} \xi_i \epsilon_{t-i}$, with $\xi_0 = 1$, to be an invertible stationary process: the sequence $\boldsymbol{\xi}_\infty := (\xi_0, \xi_1, ...)'$ is square summable and corresponds to the weights of the (purely non-deterministic) Wold-decomposition of x_t, see Brockwell and Davis (1993). Let $\boldsymbol{\Xi}$ denote the $L \cdot L$ matrix with i-th row $\boldsymbol{\Xi}_{i\cdot} := (\xi_{i-1}, \xi_{i-2}, ..., \xi_0, \mathbf{0}_{L-i})$, $i = 1, ..., L$, where $\mathbf{0}_{L-i}$ is a zero vector of length $L - i$. Define $\mathbf{x}_t := (x_t, ..., x_{t-(L-1)})'$, $\boldsymbol{\epsilon}_t := (\epsilon_t, ..., \epsilon_{t-(L-1)})'$, $\mathbf{b}_\epsilon := \boldsymbol{\Xi}\mathbf{b}_x$. Then

$$y_t = \mathbf{b}_x' \mathbf{x}_t \approx (\boldsymbol{\Xi}\mathbf{b}_x)' \boldsymbol{\epsilon}_t = \mathbf{b}_\epsilon' \boldsymbol{\epsilon}_t, \tag{27}$$

where the approximation by the finite MA inversion of x_t holds if filter coefficients decay to zero sufficiently rapidly (exact results are proposed in the appendix). The MSE predictor of $z_{t+\delta}$ is derived in McElroy and Wildi (2016)

$$\hat{\gamma}_{x\delta}(B) = \sum_{k\geq 0} \gamma_{k+\delta} B^k + \sum_{k<0} \gamma_{k+\delta} \left[\boldsymbol{\xi}(B)\right]_{|k|}^{\infty} B^k \boldsymbol{\xi}^{-1}(B), \tag{28}$$

where B is the backshift operator, $\boldsymbol{\xi}(B) = \sum_{k\geq 0} \xi_k B^k$, $\boldsymbol{\xi}^{-1}(B)$ denotes the AR-inversion and the notation $[\cdot]_{|k|}^{\infty}$ means omission of the first $|k| - 1$ lags. Intuitively, $\boldsymbol{\xi}^{-1}(B)$ transforms x_t into ϵ_t and $\gamma_{k+\delta} \left[\boldsymbol{\xi}(B)\right]_{|k|}^{\infty} B^k$ replicates the weights assigned by the target to present and past ϵ_{t-k}, $k = 0, 1, ...$ Therefore, the prediction error consists of future innovations ϵ_{t+j}, $j > 0$, and is orthogonal to the available data (MSE principle). Let $\hat{\boldsymbol{\gamma}}_{x\delta}$ denote the first L coefficients of the MSE predictor and set $\boldsymbol{\gamma}_{\Xi\delta} := \boldsymbol{\Xi}\hat{\boldsymbol{\gamma}}_{x\delta}$[7] so that $y_{MSE,t} \approx \hat{\boldsymbol{\gamma}}_{x\delta}' \mathbf{x}_t \approx \boldsymbol{\gamma}_{\Xi\delta}' \boldsymbol{\epsilon}_t$. We are then in a position to generalize Criterion (1)

$$\left.\begin{array}{c} \max_{\mathbf{b}_\epsilon} \mathbf{b}_\epsilon' \boldsymbol{\gamma}_{\Xi\delta} \\ \mathbf{b}_\epsilon' \mathbf{M} \mathbf{b}_\epsilon = \rho_1 \\ \mathbf{b}_\epsilon' \mathbf{b}_\epsilon = 1 \end{array}\right\} \tag{29}$$

assuming an arbitrary unit length or unit variance $l = 1$. The SSA solution $\mathbf{b}_x = \boldsymbol{\Xi}^{-1}\mathbf{b}_\epsilon$ is obtained by solving for $\mathbf{b}_\epsilon$ in Corollary (3), inserting $\boldsymbol{\gamma}_{\Xi\delta}$ for $\boldsymbol{\gamma}_\delta$ in Equation (4). We deduce that autocorrelation of x_t can be interpreted and treated as a modification of the target specification in the original white noise model. If y_t is (nearly) Gaussian, then $ht_1 := \pi / \arccos(\rho_1)$ expresses the HT of the predictor and the dual interpretation in Corollary (6) applies invariably. An exact expression for the SSA predictor is obtained in the appendix but in order to simplify exposition and to maintain straightforward notations, we now assume pertinence of the finite MA approximation of Equation (27). For illustration, Criterion (29) is applied to the HP target in the previous section, relying on three different AR(1) processes $x_t = a_1 x_{t-1} + \epsilon_t$ with $a_1 = -0.6, 0, 0.6$, see Figure (7).

[7] In principle we could allow for one-sided infinite expansions at this point but in applications the finite sample approximation is sufficiently accurate if L is large enough.

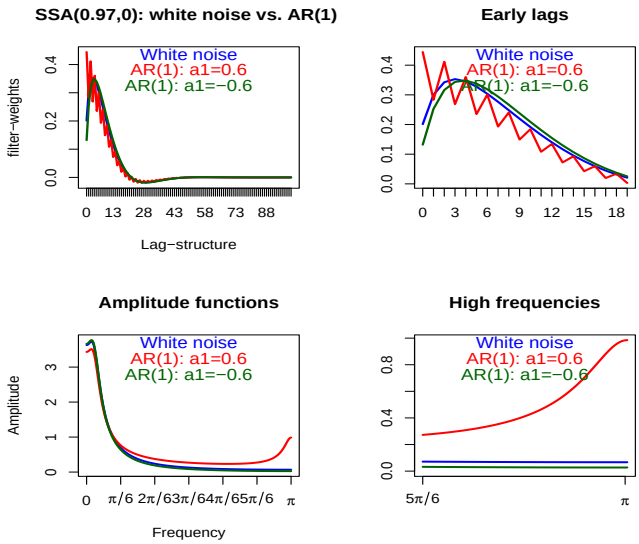

Figure 7: SSA(0.97,0) based on HP(1600)-target. Top left: filters applied to white noise (blue) and AR(1) (red and green); top-right: early lags; bottom-left: classic amplitude functions; bottom-right: classic amplitude towards higher frequencies. All filters are arbitrarily scaled to unit length.

	AR(1)=-0.6	AR(1)=0	AR(1)=0.6
HP MSE	4.344	8.138	14.742
SSA(0.97,0)	12.793	12.793	12.793

Table 5: HTs of HP (MSE predictor) and SSA as applied to three different AR(1) processes. SSA maintains a fixed HT across processes.

Table (5) reports HTs of MSE and SSA predictors: while the former depend on the data generating process (DGP), increasing markedly with a_1, the latter remain fixed, irrespective of the DGP. We deduce that the application of a fixed filter to data with unequal dependence structure can lead to qualitatively different components, for example trends or cycles in the case of HP, and SSA can address that ambiguity. For the first two processes in the first two columns of Table (5), the HTs of HP are smaller than the SSA-specification $ht = 12.79$ and SSA must increase smoothness over the benchmark. In contrast, the HT $ht = 14.74$ of the benchmark for the third process exceeds the SSA-specification and the latter 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.(7). In the frequency domain, the tail behavior of the (classic) 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.

To conclude, note that the forecast problem in Section (5.1) based on the MA(3) $z_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$ process can be expressed alternatively by assigning the role of the dependency structure of the target to the MA-inversion of x_t, i.e., $x_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$, $\xi_0 = \xi_1 = \xi_3 = 1$, $z_{t+\delta} = x_{t+\delta}$ and $\gamma_k = \begin{cases} 1 & k = 0 \\ 0 & \text{otherwise} \end{cases}$ which is a more natural or common problem formulation. In this context, the main distinguishing feature of forecasting is that the target filter $\gamma_{k+\delta}$ is an acausal identity or an anticipative *allpass* filter. In contrast, signal extraction is about the isolation of specific components in terms of acausal lowpass (trend: HP), bandpass (cycle: BK filter), multi-

trough (seasonal adjustment) or highpass (HP-gap: identity minus HP-trend). Finally, smoothing is about *causal* targets, see Section (7). We now propose an extension of the SSA framework to non-stationary integrated processes.

6.2 Integrated Processes: Emphasizing Monotonicity

The main modification of the original Criterion (1) for stationary processes concerned the target specification in the objective of Criterion (29), based on the MSE predictor in Equation (28). We now consider an extension to non-stationary integrated processes. Wildi (2024) discusses the case of a non-stationary integrated $\tilde{x}_t$ but a stationary (zero-mean) z_t, i.e., the target filter $\boldsymbol{\gamma}$ cancels the unit root(s) of $\tilde{x}_t$ such that target correlation, lag-one ACF and zero-crossings of the predictor y_t of $z_{t+\delta}$ remain well defined. This problem can be addressed in a stationary framework, after suitable transformation and is therefore omitted here. Instead, we consider non-stationary $\tilde{x}_t$, z_t and y_t so that neither the target correlation nor the rate of zero-crossings of the predictor are properly defined anymore. Let then $\tilde{x}_t$ be such that $\Delta(B)\tilde{x}_t = (1-B)^d\tilde{x}_t =: x_t$ is stationary, invertible, with Wold decomposition (weights) ξ_k, $k \geq 0$, and assume $\sum_{|k|<\infty}|\gamma_k k^d| < \infty$, which holds in typical (ARIMA-based) applications. McElroy and Wildi (2016) derive the (semi-infinite) MSE predictor $y_{t,MSE}$ with weights $\hat{\boldsymbol{\gamma}}_{\tilde{x}\delta} = (\hat{\gamma}_{\tilde{x}\delta,k})_{k\geq 0}$:

$$\hat{\gamma}_{\tilde{x}\delta}(B) = \sum_{k\geq 0}\gamma_{k+\delta}B^k + \sum_{k<0}\gamma_{k+\delta}\left(\sum_{j=1}^{d}A_{j,d+|k|}B^{d-j} + \sum_{j=1}^{|k|}\psi_{|k|-j}\left[\boldsymbol{\xi}(B)\right]_j^{\infty}B^{-j}\boldsymbol{\xi}^{-1}(B)\Delta(B)\right), \quad (30)$$

see the appendix for definition and further explanation of the weights involved. Under suitable assumptions (which typically hold for an ARIMA-based approach) $\sum_{k=0}^{\infty}|\hat{\gamma}_{\tilde{x}\delta,k}k|^d < \infty$ and

$$\sum_{k=-\infty}^{\infty}(\gamma_{k+\delta} - \hat{\gamma}_{\tilde{x}\delta,k})k^j = 0,\ j = 0, ..., d-1, \quad (31)$$

where we assume $\hat{\gamma}_{\tilde{x}\delta,k} = 0$ for $k < 0$. The constraints are necessary to cancel the unit root(s) of $\tilde{x}_t$ by the error filter $\gamma_{k+\delta} - \hat{\gamma}_{\tilde{x}\delta,k}$ with output $e_t := z_{t+\delta} - y_{t,MSE}$ such that e_t is a finite variance stationary process, as a result of MSE optimization. Equation (31) imposes cointegration between the non-stationary target $z_{t+\delta}$ and the optimal predictor and we refer to it in terms of *cointegration constraint(s)*. We now assume that similar constraints apply to the SSA predictor $\mathbf{b}_{\tilde{x}}$, see below for a formal implementation.

To simplify discussion, we here focus on the case $d = 1$ so that $\tilde{x}_t$ and z_t are both I(1) processes and therefore the cointegration constraint becomes $\sum_{k=-\infty}^{\infty}(\gamma_{k+\delta} - \hat{\gamma}_{\tilde{x}\delta,k}) = \sum_{k=-\infty}^{\infty}(\gamma_{k+\delta} - b_{\tilde{x}k}) = 0$ (a generalization to the case $d > 1$ relies on similar arguments, see the appendix for details). Let us introduce some additional notation:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ ... & & & & \\ 1 & 1 & 1 & ... & 1 \end{pmatrix} \text{ and } \boldsymbol{\Delta} := \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 1 & -1 & 0 & ... & \\ 0 & 1 & -1 & ... & 0 \\ & & & & \\ 0 & 0 & 0 & ... & 1 \end{pmatrix}$$

are the $L \cdot L$ dimensional summation and differentiation matrices, $\tilde{\mathbf{x}}_t := (\tilde{x}_t, ..., \tilde{x}_{t-(L-1)})'$ and let $\hat{\boldsymbol{\gamma}}_{MSE,L}$ with weights $(\hat{\gamma}_{MSE,Lk})_{0\leq k\leq L-1} := \kappa(\hat{\gamma}_{\tilde{x}\delta,k})_{0\leq k\leq L-1}$ denote the truncated MSE-predictor of length L, where κ is chosen such that the cointegration constraint $\sum_{k=0}^{L-1}\hat{\gamma}_{MSE,Lk} = \sum_{k=-\infty}^{\infty}\gamma_{k+\delta}$ holds[8]. Let $y_{t,MSE,L}$ designate the output of $\hat{\boldsymbol{\gamma}}_{MSE,L}$. Then

$$y_{t,MSE,L} - y_t = (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\tilde{\mathbf{x}}_t = (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\boldsymbol{\Delta}'\tilde{\mathbf{x}}_t = (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\mathbf{x}_t. \quad (32)$$

[8] Alternatively, one could alter the first weight only, assigned to lag zero. In any case, the transformation becomes an identity as $L \to \infty$

The last equality holds because the last entry of $(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'$ vanishes due to the cointegration constraint $\sum_{k=0}^{L-1}(\hat{\gamma}_{MSE,Lk}-b_{\tilde{x}k})=0$. Therefore, the last entry $\tilde{x}_{t-(L-1)}$ of $\boldsymbol{\Delta}'\tilde{\mathbf{x}}_t$ on the left side of the equation can be replaced by $x_{t-(L-1)}=\tilde{x}_{t-(L-1)}-\tilde{x}_{t-L}$ on the right side. We now show that $(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'$ is absolutely summable, as $L\to\infty$. First, let $\alpha_k:=(\hat{\gamma}_{MSE,Lk}-b_{\tilde{x}k})$ so that $\boldsymbol{\Sigma}\boldsymbol{\alpha}=(\sum_{j=0}^{k}\alpha_j)_{k=0,\dots,L-1}=(\sum_{j>k}\alpha_j)_{k=0,\dots,L-1}$, by the cointegration constraint. Then the claim of (asymptotic) absolute summability follows from $\sum_k\left|\sum_{j=0}^{k}\alpha_j\right|=\sum_k\left|\sum_{j>k}\alpha_j\right|\leq\sum_k\sum_{j\geq k}|\alpha_j|=\sum_k|k\alpha_k|\leq\sum_k|k\hat{\gamma}_{MSE,Lk}|+\sum_k|kb_{\tilde{x}k}|$ which is bounded as a function of L[9]. We then deduce that the finite MA inversion $(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\mathbf{x}_t\approx(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t$ holds and that the error is negligible if the MA-inversion $\xi_k,k\geq 0$ of x_t decays sufficiently fast for given L (exact solutions could be obtained by the proceeding outlined in the appendix but are not explicitly discussed here). We then obtain

$$y_{t,MSE,L}-y_t=(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\mathbf{x}_t\approx(\hat{\boldsymbol{\gamma}}_{MSE,L}-\mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}'\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t=\hat{\boldsymbol{\gamma}}'_{MSE,L}\boldsymbol{\Sigma}'\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t-\mathbf{b}'_{\epsilon}\boldsymbol{\Sigma}'\boldsymbol{\epsilon}_t,\quad(33)$$

where $\mathbf{b}_{\epsilon}:=\boldsymbol{\Xi}\mathbf{b}_{\tilde{x}}$ inherits the cointegration constraint from $\mathbf{b}_{\tilde{x}}$ and where we used commutativity of summation $\boldsymbol{\Sigma}$ and MA-inversion $\boldsymbol{\Xi}$ in deriving the last equality. The leftmost processes, being the predictors of $z_{t+\delta}$, are non-stationary and cointegrated; the 'synthetic' processes to the right, that is the corresponding MSE and SSA filter outputs, are *stationary* and therefore they are not predictors of the target anymore; however, by cointegration all cross-sectional differences are (approximatively) the same. Furthermore, the variances of the synthetic MSE and SSA filter outputs eventually diverge with increasing L, in contrast to the variances of the cross-sectional differences which remain bounded. Since $\mathbf{b}'_{\epsilon}\boldsymbol{\epsilon}_t=\mathbf{b}'_{\tilde{x}}\boldsymbol{\Delta}'\boldsymbol{\Sigma}'\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t\approx y_t-y_{t-1}$, we can impose a HT constraint based on the lag-one ACF of $\mathbf{b}'_{\epsilon}\boldsymbol{\epsilon}_t$ which addresses the rate of zero-crossings of the stationary y_t-y_{t-1}. Finally, we can define a target correlation, due to stationarity of the (synthetic) MSE and SSA series. In summary, Equation (33) leads to the following two equivalent expressions for the generalized SSA criterion:

$$\left.\begin{array}{c}\max_{\mathbf{b}_{\epsilon}}\mathbf{b}'_{\epsilon}\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}\\ \mathbf{b}'_{\epsilon}\mathbf{M}\mathbf{b}_{\epsilon}=\rho_1\mathbf{b}'_{\epsilon}\mathbf{b}_{\epsilon}\\ \mathbf{b}'_{\epsilon}\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_{\epsilon}=l\end{array}\right\}\quad\text{or}\quad\left.\begin{array}{c}\min_{\mathbf{b}_{\epsilon}}(\boldsymbol{\gamma}_{\tilde{\Xi}\delta}-\boldsymbol{\Sigma}\mathbf{b}_{\epsilon})'(\boldsymbol{\gamma}_{\tilde{\Xi}\delta}-\boldsymbol{\Sigma}\mathbf{b}_{\epsilon})\\ \mathbf{b}'_{\epsilon}\mathbf{M}\mathbf{b}_{\epsilon}=\rho_1\mathbf{b}'_{\epsilon}\mathbf{b}_{\epsilon}\end{array}\right\},\quad(34)$$

where $\boldsymbol{\gamma}_{\tilde{\Xi}\delta}:=\boldsymbol{\Sigma}\boldsymbol{\Xi}\boldsymbol{\gamma}_{MSE,L}$, using commutativity. The criterion on the left optimizes $\mathbf{b}_{\epsilon}$ subject to a modified length constraint $\mathbf{b}'_{\epsilon}\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_{\epsilon}=l$ that warrants proportionality of objective function and target correlation. The second criterion on the right emphasizes a MSE objective and therefore the length constraint can be skipped. For the left-hand optimization, the derivative of the Lagrangian heads towards a system of equations for $\mathbf{b}_{\epsilon}$, namely $\boldsymbol{\mathcal{V}}^{-1}\mathbf{b}_{\epsilon}=D\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$ with $\boldsymbol{\mathcal{V}}:=2\mathbf{M}-2\rho_1\mathbf{I}+\kappa\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$, $\kappa=2\tilde{\lambda}_1/\tilde{\lambda}_2$ and $D=1/\tilde{\lambda}_2$, subject to the cointegration constraint. In this framework, κ can be selected for compliance with the HT constraint and D is a scaling that ensures compliance with the length constraint (its sign ensures the positiveness of the objective function). Finally, $\mathbf{b}_{\tilde{x}}$ can be obtained from $\boldsymbol{\Xi}^{-1}\mathbf{b}_{\epsilon}$. A similar layout applies to the right-hand Criterion (34), for which the derivative of the objective function becomes $-2\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}+2\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_{\epsilon}$ leading to the Lagrangian equations $\left(-\tilde{\lambda}\left(\mathbf{M}-\rho_1\mathbf{I}\right)+\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\right)\mathbf{b}_{\epsilon}=\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$ such that $\tilde{\boldsymbol{\mathcal{V}}}^{-1}\mathbf{b}_{\epsilon}=F\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$ with $\tilde{\boldsymbol{\mathcal{V}}}:=2\mathbf{M}-2\rho_1\mathbf{I}-2/\tilde{\lambda}\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$ and $F:=-\frac{2}{\tilde{\lambda}}$. Setting $-2/\tilde{\lambda}=\kappa$ replicates the left-hand side criterion but the previously arbitrary scaling is now fixed to minimize MSE. For a (nearly) Gaussian predictor y_t, $ht_1:=\pi/\arccos(\rho_1)$ expresses the HT of y_t-y_{t-1}: interpreted in its dual form, y_t is 'most monotonic' in the sense that sign changes of y_t-y_{t-1} are fewest possible for a given tracking accuracy.

In the above derivations we assumed validity of the cointegration constraint which is now set-up explicitly to obtain the SSA predictor from the proposed criteria. Denote $\Gamma(0):=\sum_{|k|<\infty}\gamma_k$[10]

[9] The difference Equation (9) together with the boundary constraints $b_{\tilde{x},-1}=b_{\tilde{x},L}=0$ can be used to show that $\sum_k|kb_{\tilde{x}k}|$ is bounded as a function of L if $\sum_{k=0}^{\infty}|\hat{\gamma}_{\tilde{x}\delta,k}k|^d<\infty$.

[10] The notation $\Gamma(0)$ refers to the transfer function of the filter at frequency zero, see McElroy and Wildi (2016).

so that the constraint $\Gamma(0) = \sum_{k=0}^{L-1} b_{\tilde{x}k}$ can be expressed in vector notation $\mathbf{b}_{\tilde{x}} = \Gamma(0)\mathbf{e}_1 + \mathbf{B}\tilde{\mathbf{b}}$, where $\mathbf{B} = \begin{pmatrix} -1 & -1 & -1 & ... & -1 \\ 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & & & & \\ 0 & 0 & 0 & .... & 1 \end{pmatrix}$ is an $L \cdot (L-1)$ dimensional matrix, whose first row, filled with -1, is stacked on the $(L-1)\cdot(L-1)$ identity, and where the unit vector $\mathbf{e}_1 = (1, 0, ..., 0)'$ and $\tilde{\mathbf{b}} = (\tilde{b}_1, ..., \tilde{b}_{L-1})'$ are of length L and $L-1$, respectively. We then obtain

$$\mathbf{b}_\epsilon' \mathbf{M} \mathbf{b}_\epsilon = \mathbf{b}_{\tilde{x}}' \mathbf{\Xi}' \mathbf{M} \mathbf{\Xi} \mathbf{b}_{\tilde{x}} = \Gamma(0)^2 \mathbf{e}_1' \mathbf{\Xi}' \mathbf{M} \mathbf{\Xi} \mathbf{e}_1 + 2\Gamma(0) \mathbf{e}_1' \mathbf{\Xi}' \mathbf{M} \mathbf{\Xi} \mathbf{B} \tilde{\mathbf{b}} + \tilde{\mathbf{b}}' \mathbf{B}' \mathbf{\Xi}' \mathbf{M} \mathbf{\Xi} \mathbf{B} \tilde{\mathbf{b}}.$$

A similar expression is obtained for $\mathbf{b}_\epsilon' \mathbf{b}_\epsilon$, replacing $\mathbf{M}$ by $\mathbf{I}$ in the above expression. Then, the HT constraint of (either) Criterion (34) becomes

$$\Gamma(0)^2 \mathbf{e}_1' \mathbf{\Xi}' \mathcal{V}_{\rho_1} \mathbf{\Xi} \mathbf{e}_1 + 2\Gamma(0) \mathbf{e}_1' \mathbf{\Xi}' \mathcal{V}_{\rho_1} \mathbf{\Xi} \mathbf{B} \tilde{\mathbf{b}} + \tilde{\mathbf{b}}' \mathbf{B}' \mathbf{\Xi}' \mathcal{V}_{\rho_1} \mathbf{\Xi} \mathbf{B} \tilde{\mathbf{b}} = \mathbf{0},$$

where $\mathcal{V}_{\rho_1} = \mathbf{M} - \rho_1 \mathbf{I}$. Taking derivatives with respect to $\tilde{\mathbf{b}}$ gives $2\Gamma(0)\mathbf{e}_1' \mathbf{\Xi}' \mathcal{V}_{\rho_1} \mathbf{\Xi} \mathbf{B} + 2\mathbf{B}' \mathbf{\Xi}' \mathcal{V}_{\rho_1} \mathbf{\Xi} \mathbf{B} \tilde{\mathbf{b}}$. For the objective function we can rely on the right hand (MSE-) SSA criterion, thus skipping the length constraint:

$$\begin{aligned} (\boldsymbol{\gamma}_{\Xi\delta} - \mathbf{\Sigma}\mathbf{b}_\epsilon)'(\boldsymbol{\gamma}_{\Xi\delta} - \mathbf{\Sigma}\mathbf{b}_\epsilon) &= (\boldsymbol{\gamma}_{\Xi\delta} - \tilde{\mathbf{\Xi}}\mathbf{b}_{\tilde{x}})'(\boldsymbol{\gamma}_{\Xi\delta} - \tilde{\mathbf{\Xi}}\mathbf{b}_{\tilde{x}}) \\ &= \left(\boldsymbol{\gamma}_{\Xi\delta} - \tilde{\mathbf{\Xi}}\left[\Gamma(0)\mathbf{e}_1 + \mathbf{B}\tilde{\mathbf{b}}\right]\right)' \left(\boldsymbol{\gamma}_{\Xi\delta} - \tilde{\mathbf{\Xi}}\left[\Gamma(0)\mathbf{e}_1 + \mathbf{B}\tilde{\mathbf{b}}\right]\right), \end{aligned}$$

where $\tilde{\mathbf{\Xi}} := \mathbf{\Sigma}\mathbf{\Xi}$. Its derivative is $2\mathbf{B}'\tilde{\mathbf{\Xi}}'\left(\boldsymbol{\gamma}_{\Xi\delta} - \tilde{\mathbf{\Xi}}\left[\Gamma(0)\mathbf{e}_1 + \mathbf{B}\tilde{\mathbf{b}}\right]\right) = 2\mathbf{B}'\tilde{\mathbf{\Xi}}'\left(\boldsymbol{\gamma}_{\Xi\delta} - \Gamma(0)\tilde{\mathbf{\Xi}}\mathbf{e}_1\right) - 2\mathbf{B}'\tilde{\mathbf{\Xi}}'\tilde{\mathbf{\Xi}}\mathbf{B}\tilde{\mathbf{b}}$. Plugging both derivatives into the Lagrangian and equating to zero heads towards a system of equations for $\tilde{\mathbf{b}}$

$$\tilde{\mathbf{b}} = \left(\mathbf{B}'\tilde{\mathbf{\Xi}}'\tilde{\mathbf{\Xi}}\mathbf{B} + \tilde{\lambda}\mathbf{B}'\mathbf{\Xi}'\mathcal{V}_{\rho_1}\mathbf{\Xi}\mathbf{B}\right)^{-1} \left\{\mathbf{B}'\tilde{\mathbf{\Xi}}'\left(\boldsymbol{\gamma}_{\Xi\delta} - \Gamma(0)\tilde{\mathbf{\Xi}}\mathbf{e}_1\right) - \tilde{\lambda}\Gamma(0)\mathbf{e}_1'\mathbf{\Xi}'\mathcal{V}_{\rho_1}\mathbf{\Xi}\mathbf{B}\right\}. \tag{35}$$

The solution to the (right hand MSE) SSA Criterion (34), subject to the cointegration constraint reparameterized in terms of $\tilde{\mathbf{b}}$, is then obtained from Equation (35), whereby $\tilde{\lambda}$ must be selected for compliance with the HT constraint: $\mathbf{b}_\epsilon' \mathbf{M} \mathbf{b}_\epsilon = \rho_1 \mathbf{b}_\epsilon' \mathbf{b}_\epsilon$ where $\mathbf{b}_\epsilon = \mathbf{\Xi}\left(\Gamma(0)\mathbf{e}_1 + \mathbf{B}\tilde{\mathbf{b}}\right)$. Extensions to higher order integration orders $d > 1$ can be obtained by using the summation operator $\mathbf{\Sigma}^d$ in the above expressions, assuming γ_k to be such that $\sum_{|k|<\infty} |\gamma_k k^d| < \infty$ and by imposing additional (cointegration) constraints of the form $\sum_{k=0}^{L-1} b_{\tilde{x}k} k^j = \sum_{k=-\infty}^{\infty} \gamma_k k^j$, $j = 0, ..., d-1$, canceling the higher order unit root of $\tilde{x}_t$ in the prediction error $e_t = z_{t+\delta} - y_t$. An explicit extension to the practically relevant I(2)-case can be found in the appendix.

7 Smoothing: SSA vs. Whittaker-Henderson and Hodrick Prescott

An interesting simple SSA problem is obtained when selecting $\boldsymbol{\gamma} = 1$ the identity and $\delta = -(L-1)/2$ (backcast), assuming L to be an odd number. In this case, Criterion (2) becomes

$$\left.\begin{array}{c} \max_{\mathbf{b}} \rho\Big(y, x, \delta = -(L-1)/2\Big) \\ \rho(y, y, 1) = \rho_1 \end{array}\right\} \tag{36}$$

and its solution y_t 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 to Criterion (36) shifted forward and centered at x_t. We can contrast our approach with classic Whittaker-Henderson (WH) smoothing,

Whittaker (1922) and Henderson (1924), who propose to solve the following optimization problem for $\mathbf{u} := (u_1, ..., u_T)$

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

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 (36) is more apt at controlling this particular characteristic. For illustration, Fig.(8) displays HP and two different SSA smoothers, all based on a fixed length $L = 201$, and Table (6) compares their performances. For an identical holding time, SSA1 (blue

	HP	SSA1	SSA2
Holding times	34.366	34.366	42.830
Target correlations	0.271	0.301	0.271
RMS second-order differences	0.014	0.053	0.039

Table 6: 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.

line in the figure) outperforms HP in terms of target correlation. In virtue of Corollary (6), SSA2 (violet line in the figure) outperforms HP according to the holding time for given identical target correlation. However, HP 'wins' with respect to mean-square second-order differences, as a result of optimization. The observed discrepancies in each one of the reported performance measures seem sufficiently important to ask for an informed decision: minimizing the rate of zero-crossings, by SSA, or minimizing the curvature, by WH.

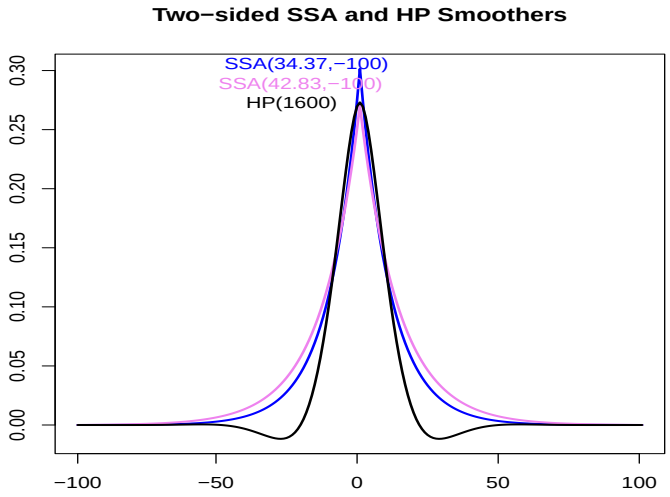

Figure 8: Coefficients of two-sided SSA and HP smoothers of length 201, arbitrarily scaled to unit length (unit variance assuming standardized white noise data): the first SSA design (blue line) replicates the holding time of HP, the second SSA design (violet line) replicates the tracking-ability or target correlation of HP.

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 (36). Furthermore, the original Criterion (1) or its extensions (29), (34) or (35) address more general estimation or prediction problems than the simple identity-smoothing (36), 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 alarming 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. Finally, our extension of SSA to non-stationary I(2) processes results in predictors whose paths are smoothest possible in the sense that the number of inflection points is minimized for a given tracking accuracy, see the appendix for reference.

8 Conclusion

We propose a novel SSA criterion that emphasizes sign accuracy and zero-crossings of the predictor subject to a HT constraint. The classic MSE approach is equivalent to unconstrained SSA-optimization: in the absence of a HT constraint and down to an arbitrary scaling nuisance. In its primal form, SSA 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 which translates into most monotonic and smoothest possible predictors in the case of I(1) and I(2) processes, respectively. Moreover, the SSA predictor is interpretable and appealing due to its actual simplicity and because the underlying criterion merges relevant concepts of prediction in terms of sign accuracy, MSE, and smoothing requirements which can be merged with timeliness, i.e. retardation or advancement, to constitute a prediction trilemma. The application to real-time signal extraction, based on the HP target, illustrates that the timeliness and smoothness of real-time designs can be controlled effectively. Finally, interpretability and economic content can be transferred from the target to the SSA predictor, due to optimality of the approximation.

9 Appendix

9.1 Theoretical vs. Empirical HTs for t-Distributed Random Variables

Table (7) evaluates the effect of heavy tails on HTs of the SSA nowcasts of Section (3), based on t-distributed white noise sequences for a broad range of degrees of freedom in $\{2.1, 4, 6, 8, 10, 100\}$ corresponding to finite variance processes.

	MSE	SSA(0.97,-100)	SSA(0.8,-100)
t-dist.: df=2.1	9.9	14.1	6.0
t-dist.: df=4	8.9	13.3	5.3
t-dist.: df=6	8.5	13.1	5.1
t-dist.: df=8	8.4	13.0	5.0
t-dist.: df=10	8.3	12.9	5.0
t-dist.: df=100	8.2	12.8	4.9
Gaussian	8.1	12.8	4.9
Theoretical HT	8.1	12.8	4.9

Table 7: The effect of heavy tails on the empirical HTs of HP predictors, based on samples of length one Million: Gaussian vs. t-distributed data and theoretical HTs.

Heavier tails increase the positive HT bias because extreme observations can trigger the impulse response of the filter, which does not change signs frequently. On the other hand, the central limit effect works against this bias in the sense that stronger smoothing of the non-Gaussian noise by the filter can narrow the gap to the Gaussian distribution; as an example, the filter in the second column (strongest smoothing) seems least affected by distortions of the HT, in relative terms,

followed by MSE and the 'unsmoothing' design in the third column. An extension to (conditional) heteroscedastic processes is discussed in Wildi (2024) who illustrates that SA and HT are fairly robust against vola-clustering.

9.2 Spherical Length- and Hyperbolic HT Constraints

Let $L \geq 3$ as assumed by Theorem (1) and consider the spectral decomposition

$$\mathbf{b} := \sum_{i=1}^{L} \alpha_i \mathbf{v}_i, \tag{37}$$

where $\sum_{i=1}^{L} \alpha_i^2 = 1$ (unit-sphere constraint, i.e., we assume $l = 1$ in the length constraint). Moreover, $\rho_1 = \mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{i=1}^{L} \alpha_i^2 \lambda_i$ so that $\alpha_{j_0} = \pm\sqrt{\frac{\rho_1}{\lambda_{j_0}} - \sum_{k \neq j_0} \alpha_k^2 \frac{\lambda_k}{\lambda_{j_0}}}$, where j_0 is such that $\lambda_{j_0} \neq 0$. The SSA problem can be solved if the hyperbola, defined by the HT constraint, intersects the unit-sphere. Plugging the former into the latter we obtain

$$\alpha_{i_0}^2 = 1 - \sum_{i \neq i_0} \alpha_i^2 = 1 - \left(\frac{\rho_1}{\lambda_{j_0}} - \sum_{k \neq j_0} \alpha_k^2 \frac{\lambda_k}{\lambda_{j_0}} \right) - \sum_{i \neq i_0, j_0} \alpha_i^2$$

for $i_0 \neq j_0$. Solving for α_{i_0} then leads to

$$\alpha_{i_0} = \pm\sqrt{\frac{\lambda_{j_0} - \rho_1}{\lambda_{j_0} - \lambda_{i_0}} - \sum_{k \neq i_0, k \neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}}}. \tag{38}$$

We now examine the case $\rho_1 = -\rho_{max}(L) = \lambda_L$ and set $i_0 = L$ so that $\rho_1 = \lambda_L$:

$$\alpha_L = \pm\sqrt{1 - \sum_{k \neq L, k \neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_L}}. \tag{39}$$

For $j_0 = L - 1$ we have $\lambda_{L-1} - \lambda_k < 0$ in the numerators of the summands in Equation 39 and $\lambda_{L-1} - \lambda_L > 0$ in the denominators. We then deduce that if $\alpha_k \neq 0$ for some $k < L - 1$, then $|\alpha_L| > 1$ which would contradict the unit-sphere constraint. Therefore, $\alpha_k = 0$ for $k < L - 1$ so that $\alpha_L = \pm 1$, $\alpha_{L-1} = 0$ and $\mathbf{b} := \pm\mathbf{v}_L$ (the contacts of unit-sphere and hyperbola are tangential at the vertices $\pm\mathbf{v}_L$). Since $w_L \neq 0$ (completeness assumption), the SSA solution $\mathbf{b} := \text{sign}(w_L)\mathbf{v}_L$ warrants a positive objective function $\boldsymbol{\gamma}_\delta'\mathbf{b} = \text{sign}(w_L)w_L > 0$, confirming Corollary (1). Next, for $\rho_1 > \lambda_L$ the quotient $\frac{\lambda_{L-1} - \rho_1}{\lambda_{L-1} - \lambda_L}$ in Equation (38) (still assuming $j_0 = L - 1$) is smaller one which allows for non-vanishing $\alpha_k \neq 0$, $k < L - 1$, in Equation (39). However, in this case the number under the square root should remain positive which is always the case for $\rho_1 \leq \rho_{max}(L) = \lambda_1$ since the term $-\alpha_1^2 \frac{\lambda_{L-1} - \lambda_1}{\lambda_{L-1} - \lambda_L}$ of the sum $-\sum_{k<L-1} \alpha_k^2 \frac{\lambda_{L-1} - \lambda_k}{\lambda_{L-1} - \lambda_L}$ can compensate for a potentially negative value of $\frac{\lambda_{L-1} - \rho_1}{\lambda_{L-1} - \lambda_L}$. In particular, if $\rho_1 = \rho_{max}(L)$, then positiveness of the term under the square root implies $\alpha_1 = 1$ and therefore $\alpha_2 = ... = \alpha_L = 0$, by the length constraint so that $\mathbf{b} = \pm\mathbf{v}_1$, confirming again Corollary (1). In between, that is for $\rho_{min}(L) < \rho_1 < \rho_{max}(L)$, the number under the square root of Equation (39) is in the open unit interval $]0, 1[$ and the intersection of the unit sphere and HT hyperbola is non-empty and of dimension $L - 2 \geq 1$.

9.3 The Choice of M

The matrix $\mathbf{M}$ is not uniquely determined by the quadratic form $\mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{k=1}^{L-1} b_{k-1} b_k$. Indeed, the family of corresponding matrices $\mathbf{M}(\kappa)$ has zeroes everywhere except above and below the

main diagonal, where the corresponding entries are κ and $1-\kappa$, respectively

$$\mathbf{M}(\kappa) = \begin{pmatrix} 0 & \kappa & 0 & 0 & 0 & ... & 0 & 0 & 0 \\ 1-\kappa & 0 & \kappa & 0 & 0 & ... & 0 & 0 & 0 \\ ... & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & ... & 1-\kappa & 0 & \kappa \\ 0 & 0 & 0 & 0 & 0 & ... & 0 & 1-\kappa & 0 \end{pmatrix},$$

with $\kappa \in \mathbb{R}$: for $\kappa = 0.5$ the original $\mathbf{M} := \mathbf{M}(0.5)$ is obtained. Evidently, all obtained results straightforwardly extend to $\mathbf{M}(\kappa)$, noting that derivatives of quadratic forms would involve $\mathbf{M}(\kappa) + \mathbf{M}(\kappa)' = 2\mathbf{M}(0.5)$ so that κ is canceled. More generally, the technical elements of the proof of Theorem (1) assume $\mathbf{M}$ to be symmetric with pairwise different eigenvalues and therefore the theorem could be extended to constraints of the ACF at higher lags, possibly involving the spectral density of the process, though this possibility is neither explicitly required nor explored here.

9.4 Proofs of Corollaries (2) and (4) and an Illustration of Incomplete Spectral Support

Proof of Corollary (2): The first assertion follows directly from the Lagrangian Equation (8). Under the case posited in the second assertion, $\mathbf{N}_{i_0}$ is not of full rank and $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ as defined by Equation (17) is a solution to the Lagrangian equation $D\boldsymbol{\gamma}_\delta = \mathbf{N}_{i_0}\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ for arbitrary $\tilde{N}_{i_0}$. Moreover,

$$\rho_{i_0}(\tilde{N}_{i_0}) := \frac{\mathbf{b}_{i_0}(\tilde{N}_{i_0})'\mathbf{M}\mathbf{b}_{i_0}(\tilde{N}_{i_0})}{\mathbf{b}'_{i_0}(\tilde{N}_{i_0})\mathbf{b}_{i_0}(\tilde{N}_{i_0})} = \frac{\sum_{i\neq i_0}\lambda_i w_i^2 \frac{1}{(2\lambda_i-\nu)^2} + \tilde{N}_{i_0}^2\lambda_{i_0}}{\sum_{i\neq i_0} w_i^2 \frac{1}{(2\lambda_i-\nu)^2} + \tilde{N}_{i_0}^2} = \frac{M_{i_0 1} + \tilde{N}_{i_0}^2\lambda_{i_0}}{M_{i_0 2} + \tilde{N}_{i_0}^2}.$$

Solving for the HT constraint $\rho_{i_0}(\tilde{N}_{i_0}) = \rho_1$ then leads to $\tilde{N}_{i_0}^2 = \frac{\rho_1 M_{i_0 2} - M_{i_0 1}}{\lambda_{i_0} - \rho_1}$. We infer that $\tilde{N}_{i_0}^2$ is positive if $0 < \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} < \rho_1 < \lambda_{i_0}$ or $0 > \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} > \rho_1 > \lambda_{i_0}$, as claimed. Finally, the correct sign combination of the pair $D, \tilde{N}_{i_0}$ is determined by the maximal criterion value. For a proof of the third and last assertion, we first assume that $\boldsymbol{\gamma}_\delta$ is not band-limited so that $w_1 \neq 0$ and $w_L \neq 0$. Then, $\lim_{\nu\to 2\lambda_1}\rho(\nu) = \lambda_1 = \rho_{max}(L)$ and $\lim_{\nu\to 2\lambda_L}\rho(\nu) = \lambda_L = -\rho_{max}(L)$, see the proof of Theorem (1). By continuity of $\rho(\nu)$ and by the intermediate-value theorem, any ρ_1 such that $|\rho_1| \leq \rho_{max}(L)$ is admissible for the HT constraint. Otherwise, if $w_1 = 0$ then $\mathbf{b}_1(\tilde{N}_1)$, where $i_0 = 1$ in Equation (17), can 'fill the gap' and approach the upper boundary $\rho_{max}(L)$ arbitrarily closely as $\tilde{N}_1$ increases. Note, however, that $\lim_{|\tilde{N}_1|\to\infty}\mathbf{b}_1(\tilde{N}_1) \propto \mathbf{v}_1$ would not correlate with the target anymore, so that a proper solution to the SSA problem would not exist, and therefore we must require $\rho_1 < \rho_{max}(L)$ in this case, as claimed. A similar reasoning applies if $w_L = 0$, such that $\rho_1 > -\rho_{max}(L)$. □

To illustrate the case of incomplete spectral support addressed by Corollary (2), we propose a simple nowcast example, $\delta = 0$, based on a band-limited target $\boldsymbol{\gamma}_0 = \sum_{i=4}^{10} 0.378\mathbf{v}_i$ of length $L = 10$, where $\mathbf{v}_i$ are the eigenvectors of the $10 \cdot 10$-dimensional $\mathbf{M}$, assuming that the first three weights $w_1 = w_2 = w_3 = 0$ vanish ($n = 4$ in Equation (3)) and that the last seven weights are constant $w_i = 0.378$: this particular weighting scheme implies that $\boldsymbol{\gamma}_0'\boldsymbol{\gamma}_0 = 1$ such that the SSA objective function is also the target correlation. The left panel in Fig. 9 displays the lag-one ACF 16 of $\mathbf{b}(\nu)$ given by 15 as a function of $\nu \in [-2, 2] - \{2\lambda_i, i = 1, ..., L\}$, thus omitting all potential singularities at $\nu = 2\lambda_i$, $i = 1, ..., L$; the right panel displays additionally the lag-one ACF 18 of the extension $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ in 17, when $\nu = \nu_{i_0} = 2\lambda_{i_0}$ for $i_0 = 1, 2, 3$, where the three additional (vertical black) spectral lines, corresponding to $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, show the range of ACF-values as a function of $\tilde{N}_{i_0} \in \mathbb{R}$: lower and upper bounds of each spectral line correspond to $\rho_{i_0}(0) = \rho_{\nu_{i_0}} = \frac{M_{i_0 1}}{M_{i_0 2}}$, when $\tilde{N}_{i_0} = 0$ in 18, and $\rho_{i_0}(\pm\infty) = \lambda_{i_0}$, when $\tilde{N}_{i_0} = \pm\infty$. The green horizontal lines in both graphs correspond to two different arbitrary holding-times $\rho_1 = 0.6$ and $\rho_1 = 0.365$: the intersections of the latter with the ACF, marked by colored vertical lines in each panel, indicate potential solutions

to the SSA problem for the thusly specified HT constraint. The corresponding criterion values are reported at the bottom of the colored vertical lines: the SSA solution is determined by the intersection which leads to the highest criterion value (rightmost in this example).

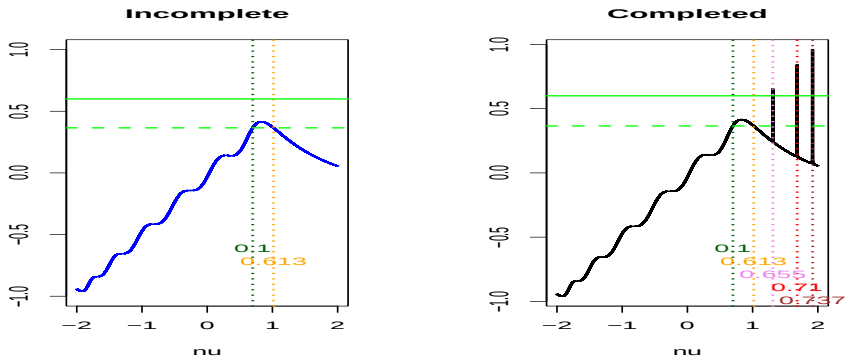

Figure 9: Lag-one autocorrelation as a function of ν. Original (incomplete) solutions (left panel) vs. completed solutions (right panel). Intersections of the ACF with the two green lines are potential solutions to the SSA problem for the corresponding HTs: criterion values are reported for each intersection (bottom right).

The right panel in the figure illustrates that the completion with the extensions $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ at the singular points $\nu = \nu_{i_0} = 2\lambda_{i_0}$ for $i_0 = 1, 2, 3$ can accommodate for a wider range of HT constraints, such that $|\rho_1| < \rho_{max}(L) = \lambda_1 = 0.959$; in contrast, $\mathbf{b}(\nu)$ in the left panel is limited to $-0.959 = \lambda_{10} < \rho_1 < \lambda_4 = 0.415$ so that there does not exist a solution for $\rho_1 = 0.6$ (no intersection with upper green line in left panel). Moreover, for a given HT constraint, the additional stationary points corresponding to intersections at the spectral lines of the (completed) ACF might lead to improved performances, as shown in the right panel, where the maximal criterion value

$$\left(\mathbf{b}_{i_0}(\tilde{N}_{i_0})\right)'\boldsymbol{\gamma}_\delta = \left(\mathbf{b}_{10}(0.077)\right)'\boldsymbol{\gamma}_0 = 0.737$$

is attained at the right-most spectral line, for $i_0 = 1$, and where $\tilde{N}_1 = 0.077$ has been obtained from 19, with the correct signs of D and $\tilde{N}_1$ in place.

To conclude we provide a **proof of Corollary** (4): let $\gamma_{k+\delta} = \lambda^{k+\delta}$. Then

$$\tilde{b}_k := D\frac{\lambda^\delta}{\lambda^2 - \nu\lambda + 1}\lambda^{k+1} \propto \lambda^{k+\delta} \tag{40}$$

is a solution to

$$\tilde{b}_{k+1} - \nu\tilde{b}_k + \tilde{b}_{k-1} \quad = \quad D\gamma_{k+\delta}\ ,\ 0 \le k \le L-1$$

with boundaries $\tilde{b}_{-1} \neq 0, \tilde{b}_L \approx 0$. The expression is well-defined because $\lambda^2 - \nu\lambda + 1 \neq 0$ since $\lambda \neq \lambda_{\rho_1}$, by assumption. Consider now the solution to the homogeneous difference-equation $\breve{b}_{k+1} - \nu\breve{b}_k + \breve{b}_{k-1} = 0$, namely $\breve{b}_k = C_1\lambda_{\rho_1}^k + C_2\lambda_{\rho_1}^{-k}$, where C_1, C_2 are arbitrary constants. We can now combine $\tilde{b}_k$ and $\breve{b}_k$ as in

$$b_k = b_k(\lambda_{1\rho_1}) \propto \lambda^{k+\delta} + C_1\lambda_{\rho_1}^k + C_2\lambda_{\rho_1}^{L-k} \approx \lambda^{k+\delta} - \lambda_{\rho_1}\lambda^{-1+\delta}\lambda_{\rho_1}^k, \tag{41}$$

where we selected $C_1 = -\lambda_{\rho_1}\lambda^{-1+\delta}$, $C_2 = 0$ such that $b_{-1} = 0$ and $\tilde{b}_L \approx 0$ (the approximation error can be neglected by the last assumption of the corollary). Corollary (3) states that the

unknown stable root λ_{ρ_1} is determined uniquely by the HT constraint

$$\frac{\sum_{k=1}^{L-1} b_k(\lambda_{\rho_1}) b_{k-1}(\lambda_{\rho_1})}{\sum_{k=0}^{L-1} b_k(\lambda_{\rho_1})^2} = \rho_1. \tag{42}$$

We can now insert 41 into this equation and solve for λ_{ρ_1}. Specifically, the numerator becomes

$$\sum_{k=1}^{L-1} b_k b_{k-1} \quad = \quad \sum_{k=1}^{L-1} \left(\lambda^{k+\delta} - \lambda_{\rho_1}\lambda^{-1+\delta}\lambda_{\rho_1}^{k}\right)\left(\lambda^{k-1+\delta} - \lambda_{\rho_1}\lambda^{-1+\delta}\lambda_{\rho_1}^{k-1}\right). \tag{43}$$

The first cross-product of terms in parantheses is

$$\lambda^{-1}\sum_{k=1}^{L-1}\lambda^{2(k+\delta)} = \lambda^{1+2\delta}\sum_{k=0}^{L-2}\lambda^{2k} = \lambda^{1+2\delta}\frac{1-\lambda^{2(L-1)}}{1-\lambda^2} \approx \frac{\lambda^{1+2\delta}}{1-\lambda^2}. \tag{44}$$

The second cross-product of terms in parentheses is

$$-\lambda_{\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k+\delta}\lambda_{\rho_1}^{k-1} = -\lambda_{\rho_1}\lambda^{2\delta}\sum_{k=0}^{L-2}(\lambda\lambda_{\rho_1})^k = -\lambda_{\rho_1}\lambda^{2\delta}\frac{1-(\lambda\lambda_{\rho_1})^{L-1}}{1-\lambda\lambda_{\rho_1}} \approx \frac{-\lambda_{\rho_1}\lambda^{2\delta}}{1-\lambda\lambda_{\rho_1}}. \tag{45}$$

The third cross-product of terms in parentheses is

$$-\lambda_{\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k-1+\delta}\lambda_{\rho_1}^{k} \approx \frac{-\lambda_{\rho_1}^2\lambda^{2\delta-1}}{1-\lambda\lambda_{\rho_1}}. \tag{46}$$

Finally, the last cross-product of terms in parantheses is

$$\lambda_{\rho_1}^2\lambda^{-2+2\delta}\sum_{k=1}^{L-1}\lambda_{\rho_1}^{2k-1} = \lambda_{\rho_1}^3\lambda^{-2+2\delta}\sum_{k=0}^{L-2}\lambda_{\rho_1}^{2k} = \lambda_{\rho_1}^3\lambda^{-2+2\delta}\frac{1-\lambda_{\rho_1}^{2(L-1)}}{1-\lambda_{\rho_1}^2} \approx \frac{\lambda_{\rho_1}^3\lambda^{-2+2\delta}}{1-\lambda_{\rho_1}^2}. \tag{47}$$

The common denominator of 44, 45, 46 and 47 is

$$(1-\lambda^2)(1-\lambda\lambda_{\rho_1})(1-\lambda_{\rho_1}^2). \tag{48}$$

Summing all terms in 44, 45, 46 and 47 under the common denominator 48 leads to a third-order polynomial

$$f_1(\lambda_{\rho_1}) := a_3\lambda_{\rho_1}^3 + a_2\lambda_{\rho_1}^2 + a_1\lambda_{\rho_1} + a_0$$

in λ_{ρ_1} with coefficients $a_3 = \lambda^{2\delta}\lambda^{-2}$, $a_2 = -\lambda^{2\delta}\lambda^{-1}$, $a_1 = -\lambda^{2\delta}$ and $a_0 = \lambda^{2\delta}\lambda$. The coefficient of order four vanishes due to cancellation of cross-terms. The same proceeding can be applied to the denominator $\sum_{k=0}^{L-1} b_k^2$ in 42:

$$\sum_{k=0}^{L-1} b_k^2 \quad = \quad \sum_{k=0}^{L-1}\left(\lambda^{k+\delta} - \lambda_{\rho_1}\lambda^{-1+\delta}\lambda_{\rho_1}^{k}\right)^2 \tag{49}$$

with cross-products of terms in parentheses

$$\begin{aligned}
\sum_{k=0}^{L-1}\lambda^{2(k+\delta)} &\approx \frac{\lambda^{2\delta}}{1-\lambda^2} \\
-2\lambda_{\rho_1}\lambda^{2\delta-1}\sum_{k=0}^{L-1}(\lambda\lambda_{\rho_1})^k &\approx -\frac{\lambda_{\rho_1}\lambda^{2\delta-1}}{1-\lambda\lambda_{\rho_1}} \\
\lambda_{\rho_1}^2\lambda^{-2+2\delta}\sum_{k=0}^{L-1}\lambda_{\rho_1}^{2k} &\approx \frac{\lambda_{\rho_1}^2\lambda^{-2+2\delta}}{1-\lambda_{\rho_1}^2}.
\end{aligned}$$

This will lead to the sum of four terms whose common denominator is 48 and whose numerator is a polynomial

$$f_2(\lambda_{\rho_1}) := b_3\lambda_{\rho_1}^3 + b_2\lambda_{\rho_1}^2 + b_1\lambda_{\rho_1} + b_0$$

with polynomial coefficients $b_3 = \lambda^{2\delta}\lambda^{-1}$, $b_2 = \lambda^{2\delta}(\lambda^{-2} - 2)$, $b_1 = \lambda^{2\delta}(\lambda - 2\lambda^{-1})$ and $b_0 = \lambda^{2\delta}$. After cancelation of their common denominator 48, equation 42 can be re-written as

$$\frac{f_1(\lambda_{\rho_1})}{f_2(\lambda_{\rho_1})} = \rho_1,$$

where the common $\lambda^{2\delta}$ of a_0, a_1, a_2, a_3 and b_0, b_1, b_2, b_3 cancel. We then obtain an equation $f_3(\lambda_{\rho_1}, \rho_1) = 0$ for λ_{ρ_1} where $f_3(\lambda_{\rho_1}, \rho_1)$ is a cubic polynomial in λ_{ρ_1} with coefficients

$$c_3 = \lambda^{-2} - \rho_1\lambda^{-1}\ ,\ c_2 = -\lambda^{-1} - \rho_1(\lambda^{-2} - 2)\ ,\ c_1 = -1 - \rho_1(\lambda - 2\lambda^{-1})\ ,\ c_0 = \lambda - \rho_1.$$

The remainder of the proof then follows from a closed-form expression for the root of a cubic polynomial (Cardano's formula). Note that $|\rho_1| < 1$ implies $c_3 \neq 0$ so that the solution λ_{ρ_1} to the HT Equation (42) is the root of a cubic polynomial, as claimed.

Remarks

In contrast to the frequency-domain decomposition 5 of the lag-one autocorrelation, the above time-domain decomposition allows for notable simplifications resulting in a closed-form solution for λ_{ρ_1} under the posited assumptions. Note also that if $|\nu| < 2$ then $\lambda_{1\rho_0}$ is a unit-root so that (some of) the geometric series in the above proof do not converge anymore and therefore high-order terms (of power L) cannot be neglected. The resulting inflated polynomial order is an indication of multiple solutions for ν given ρ_1 in this case. Finally, the solution to Equation (42) in the case of an AR(p) target with $p > 1$ would involve the root of a higher-order polynomial for which a closed-form expression does not exist in general.

9.5 Exact Solution in the Case of Autocorrelation

We derive an exact expression for the SSA predictor in the case of autocorrelated processes. To simplify notation we assume an AR(1)-process $x_t = a_1 x_{t-1} + \epsilon_t$ but our proceeding is otherwise general. In this case $x_{t-L+j} = \sum_{k=0}^{j-1} \xi_k \epsilon_{t-L+j-k} + \xi_j x_{t-L}$ with $\xi_k = a_1^k$ and therefore

$$y_t = \mathbf{b}_x' \mathbf{x}_t = (\Xi \mathbf{b}_x)' \boldsymbol{\epsilon}_t + \mathbf{b}_x' \boldsymbol{\xi}_L x_{t-L} = \mathbf{b}_\epsilon' \left(\boldsymbol{\epsilon}_t + \Xi^{-1\ \prime} \boldsymbol{\xi}_L x_{t-L}\right) = \mathbf{b}_\epsilon' \left(\boldsymbol{\epsilon}_t + a_1 \mathbf{e}_L x_{t-L}\right), \tag{50}$$

where $\boldsymbol{\xi}_L := (\xi_L, ..., \xi_1)'$, $\mathbf{e}_L = (0, ..., 0, 1)'$ is a unit vector of length L and the last equality follows from definition of Ξ^{-1} and $\boldsymbol{\xi}_L$. Factually, the approximation error in (27) originates in the replacement of $b_{xk} x_{t-k}$ by $b_{xk}(\epsilon_{t-k} + \sum_{j=1}^{L-1-k} \psi_j x_{t-k-j})$, where $x_{t-k} = \epsilon_t + \sum_{j \geq 1} \psi_j x_{t-k-j}$ denotes the AR inversion of x_{t-k}. The difference or approximation error $b_{xk} \sum_{j \geq 0} \psi_{L-k+j} x_{t-L-j}$ is attributable to those lags in the AR inversion of x_{t-k} which are not part of the filter: for an AR(1), $\psi_j = \begin{cases} a_1 & j = 1 \\ 0 & j > 1 \end{cases}$ and the rightmost expression in Equation (50) is obtained . We now refer to $\mathbf{b}_\epsilon' \boldsymbol{\epsilon}_t$ and $a_1 \mathbf{e}_L x_{t-L}$ in Equation (50) in terms of main predictor and residual, respectively, and we consider the exact MSE-predictor $y_{t,\epsilon,MSE}$ as applied to $\boldsymbol{\epsilon}_{t\infty} = (\epsilon_t, \epsilon_{t-1}, ...)'$, the semi infinite extension of $\boldsymbol{\epsilon}_t$, with weights $\boldsymbol{\gamma}_{\Xi\delta\infty} := \Xi_\infty \hat{\boldsymbol{\gamma}}_{x\delta\infty}$ based on the semi-infinite extension of Ξ applied to $\hat{\boldsymbol{\gamma}}_{x\delta\infty}$ specified by Equation (28). For a derivation of the exact SSA solution, we aim at maximizing the target correlation of y_t with $y_{t,\epsilon,MSE}$ subject to exact length and HT constraints.

Proposition 3. *Let the above assumptions about x_t and z_t hold. Then the exact finite-length SSA predictor $y_t = \mathbf{b}_x' \mathbf{x}_t = \mathbf{b}_\epsilon' (\boldsymbol{\epsilon}_t + a_1 \mathbf{e}_L x_{t-L})$ is obtained from*

$$\mathbf{b}_\epsilon(\nu_1) \quad = \quad D(\nu_1, l) \tilde{\mathbf{N}}^{-1} \left(\boldsymbol{\gamma}_{\Xi\delta 1:L} + a_1 \mathbf{e}_L \boldsymbol{\xi}_\infty' \boldsymbol{\gamma}_{\Xi\delta(L+1):\infty}\right), \tag{51}$$

where the subscripts $1:L$ *and* $(L+1):\infty$ *of* $\boldsymbol{\gamma}_{\Xi\delta}$ *signify corresponding vector entries or lags and where*

$$\tilde{\mathbf{N}} = 2\left(\mathbf{M} + a_1\left(1+\frac{a_1^2}{1-a_1^2}\right)\mathbf{e}_L\mathbf{e}_L' - \nu_1\left[\mathbf{I} + \frac{a_1^2}{1-a_1^2}\mathbf{e}_L\mathbf{e}_L'\right]\right),$$

The pairing $(\nu_1, D(\nu_1, l))$ *ensures compliance with HT and length constraints.*

Proof: Noting that $\boldsymbol{\epsilon}_t$ and x_{t-L} are uncorrelated in $y_t = \mathbf{b}_\epsilon'(\boldsymbol{\epsilon}_t + a_1\mathbf{e}_L x_{t-L})$, the length constraint (unit variance) becomes

$$\mathbf{b}_\epsilon'\left(\mathbf{I} + \frac{a_1^2}{1-a_1^2}\mathbf{e}_L\mathbf{e}_L'\right)\mathbf{b}_\epsilon = 1, \tag{52}$$

where $1/(1-a_1^2) = \sum_{k\geq 0}\xi_k^2$ is the variance or length of x_{t-L}. For the HT constraint we are looking at the lag-one ACF

$$E\left[\mathbf{b}_\epsilon'(\boldsymbol{\epsilon}_t + a_1\mathbf{e}_L x_{t-L})(\boldsymbol{\epsilon}_{t-1} + a_1\mathbf{e}_L x_{t-1-L})'\mathbf{b}_\epsilon\right],$$

acknowledging that the variance in the denominator of the ACF is unity, due to the length constraint. The above expression becomes

$$\mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon + a_1\mathbf{b}_\epsilon'\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon + \frac{a_1^3}{1-a_1^2}\mathbf{b}_\epsilon\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon, \tag{53}$$

noting that $\boldsymbol{\epsilon}_t$ and x_{t-1-L} are uncorrelated, that $a_1E[\mathbf{b}_\epsilon'\mathbf{e}_L x_{t-L}\boldsymbol{\epsilon}_{t-1}'\mathbf{b}_\epsilon] = a_1\mathbf{b}_\epsilon'\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon$ and that $a_1^2E[\mathbf{b}_\epsilon'\mathbf{e}_L x_{t-L}x_{t-1-L}\mathbf{e}_L'\mathbf{b}_\epsilon] = \frac{a_1^3}{1-a_1^2}\mathbf{b}_\epsilon'\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon$. We consider next the covariance of the SSA predictor y_t with $y_{t,\epsilon,MSE}$, splitting the task into (mutually independent) residual and main predictor of y_t. The residual $a_1\mathbf{b}_\epsilon'\mathbf{e}_L x_{t-L}$ correlates with $y_{t,\epsilon,MSE}$ by way of common ϵ_{t-L-k}, $k = 0, 1, ...$ in $x_{t-L} = \boldsymbol{\xi}_\infty'\boldsymbol{\epsilon}_{t-L\infty}$ and $\boldsymbol{\gamma}_{\Xi\delta\infty}\boldsymbol{\epsilon}_{t\infty}$. Aggregating over common terms we obtain $a_1\mathbf{b}_\epsilon'\mathbf{e}_L\boldsymbol{\xi}_\infty'\boldsymbol{\gamma}_{\Xi\delta(L+1):\infty}$ for the residual covariance. Similarly, the covariance between the main predictor $\mathbf{b}_\epsilon'\boldsymbol{\epsilon}_t$ and $\boldsymbol{\gamma}_{\Xi\delta\infty}\boldsymbol{\epsilon}_{t,\infty}$ is $\mathbf{b}_\epsilon'\boldsymbol{\gamma}_{\Xi\delta 1:L}$. Summing both contributions we obtain

$$E[y_t y_{t,\epsilon,MSE}] = \mathbf{b}_\epsilon'\left(\boldsymbol{\gamma}_{\Xi\delta 1:L} + a_1\mathbf{e}_L\boldsymbol{\xi}_\infty'\boldsymbol{\gamma}_{\Xi\delta(L+1):\infty}\right) \tag{54}$$

for the covariance of y_t and $y_{t,\epsilon,MSE}$. If the length constraint (52) is imposed, then the covariance (54) is proportional to the target correlation and therefore we can proceed to optimization, maximizing (54) subject to (52) and (53). Taking derivatives of the Lagrangian then leads to a system of equations for $\mathbf{b}_\epsilon$

$$\begin{aligned} & 2\tilde{\lambda}_2\left\{\mathbf{M} + a_1\left(1+\frac{a_1^2}{1-a_1^2}\right)\mathbf{e}_L\mathbf{e}_L'\right\}\mathbf{b}_\epsilon + 2\tilde{\lambda}_1\left[\mathbf{I} + \frac{a_1^2}{1-a_1^2}\mathbf{e}_L\mathbf{e}_L'\right]\mathbf{b}_\epsilon \\ = \;& \boldsymbol{\gamma}_{\Xi\delta 1:L} + a_1\mathbf{e}_L\boldsymbol{\xi}_\infty'\boldsymbol{\gamma}_{\Xi\delta(L+1):\infty} \end{aligned}$$

from which Equation (51) can be inferred. The resulting solution maximizes the correlation with the exact (infinite length) MSE-predictor, up to a fixed scaling constant and subject to exact lag-one and length constraints and therefore $y_t = \mathbf{b}_x'\mathbf{x}_t = \mathbf{b}_\epsilon'(\boldsymbol{\epsilon}_t + a_1\mathbf{e}_L x_{t-L})$ can be interpreted as (exact) SSA predictor, as claimed. □

Remarks: in principle, the above proof can be extended to arbitrary stationary or integrated (invertible) processes but the correction terms will be more complex than for the considered AR(1) process, thus cluttering the notation correspondingly. In general, ξ_k and ϵ_j are unknown and must be estimated: we refer to textbooks on the topic, see Brockwell and Davis (1993). Wildi (2024) shows that the impact of the finite sample estimation error remains negligible for practical sample sizes ($T = 120$ observations corresponding to 10 years of monthly data), assuming the lag-one autocorrelation is not too large; specifically, values smaller than 0.9 in absolute value help mitigate finite sample biases and outliers of the HT. In any case, we may refer to our SSA package for a quantification of corresponding effects.

9.6 MSE Predictor in the Non-Stationary Case and an Extension to I(2) Processes

McElroy and Wildi (2016) derive the semi-infinite MSE predictor $y_{t,MSE} := \hat{\gamma}_{\tilde{x}\delta}(B)\tilde{x}_t$ of a target $z_{t+\delta}$ for non-stationary $\tilde{x}_t$ such that $x_t = (1-B)^d\tilde{x}_t$ is a stationary and invertible process with Wold decomposition $x_t = \boldsymbol{\xi}(B)\epsilon_t$. Specifically, the weights $\hat{\boldsymbol{\gamma}}_{\tilde{x}\delta} = (\hat{\gamma}_{\tilde{x}\delta,0}, \hat{\gamma}_{\tilde{x}\delta,1}, ...)'$ of the predictor are

$$\hat{\gamma}_{\tilde{x}\delta}(B) = \sum_{k\geq 0}\gamma_{k+\delta}B^k + \sum_{k<0}\gamma_{k+\delta}\left(\sum_{j=1}^{d}A_{j,d+|k|}B^{d-j} + \sum_{j=1}^{|k|}\psi_{|k|-j}\left[\boldsymbol{\xi}(B)\right]_j^{\infty}B^{-j}\boldsymbol{\xi}^{-1}(B)\Delta(B)\right),$$

where $A_{jt} := \psi_{t-j} - \sum_{k=1}^{d-j}\delta_k\psi_{t-j-k}$ and where δ_k, ψ_k are obtained from $\Delta(B) = (1-B)^d =: 1 - \sum_{j=1}^{d}\delta_j B^j$, $\Psi(B) := 1/\Delta(B) = \sum_{j\geq 0}\psi_j B^j$. While the general intuition remains the same as in the stationary case of Equation (28), the additional term in $A_{j,d+|k|}$ accounts for a polynomial $p(t)$ of order $d-1$ of t in the null space of the difference operator $\Delta(B)$, i.e., a solution to the homogeneous difference equation $\Delta(B)p(t) = 0$: the coefficients of this polynomial are determined by a proper initialization of the process (boundary conditions with permanent effect). We then infer that the MSE predictor can be decomposed into a 'pure' MA inversion and $p(t)$, the latter for matching said boundary conditions. The cointegration constraints $\sum_{k=-\infty}^{\infty}(\gamma_{k+\delta} - \hat{\gamma}_{\tilde{x}\delta,k})k^j = 0$, for $j = 0, ..., d-1$, imply that the error filter $\gamma_{k+\delta} - \hat{\gamma}_{\tilde{x}\delta,k}$ cancels a time polynomial of order $d-1$ such that $p(t)$ does not appear in the prediction error $e_t = z_{t+\delta} - y_{t,MSE}$ which is a finite variance stationary process, as a result of MSE optimization.

We now briefly discuss an extension of the I(1) case discussed in Section (6.2) to I(2) processes, assuming $d = 2$. Then, the cointegration constraints can be split into a level constraint $\sum_{k=0}^{L-1}b_{\tilde{x}k} = \sum_{|k|<\infty}\gamma_k =: \Gamma(0)$ and a slope constraint $\sum_{k=1}^{L-1}kb_{\tilde{x}k} = \sum_{|k|<\infty}k\gamma_k =: \dot{\Gamma}(0)$. We denote by $y_{t,MSE,L}$, with coefficients $\hat{\boldsymbol{\gamma}}_{MSE,L}$, the truncated MSE filter, suitably modified to comply with level and slope constraints[11]. A cancellation of the double unit-root by the error filter $e_t = y_{t,MSE,L} - y_t$ is then obtained from

$$\begin{aligned} y_{t,MSE,L} - y_t &= (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\tilde{\mathbf{x}}_t = (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}^{2\,\prime}\boldsymbol{\Delta}^{2\,\prime}\tilde{\mathbf{x}}_t \\ &= (\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}^{2\,\prime}\mathbf{x}_t, \end{aligned} \tag{55}$$

where the first $L-2$ entries of $\boldsymbol{\Delta}^{2\,\prime}\tilde{\mathbf{x}}_t$ are the stationary second order differences $x_t, x_{t-1}, ..., x_{t-(L-3)}$. We also substitued the last two entries $x_{t-(L-2)}, x_{t-(L-1)}$ of $\mathbf{x}_t$ to the right of the last equation for the last two non-stationary entries of $\boldsymbol{\Delta}^{2\,\prime}\tilde{\mathbf{x}}_t$ to the left because the last two columns of $\boldsymbol{\Sigma}^{2\,\prime}$, i.e., $(L-1-k)_{k=0,...,L-1}$ and $(L-k)_{k=0,...,L-1}$, are cancelled by the cointegration constraints. Therefore, the last two entries of $(\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}^{2\,\prime}$ vanish and Equation (55) applies. In the next step, we want to replace $\mathbf{x}_t$ by the finite MA-inversion $\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t$ in Equation (55). For this purpose, we rely on boundedness of $\sum_{k=0}^{L-1}|b_{\tilde{x},k}k^2|$ and $\sum_{k=0}^{L-1}|\hat{\gamma}_{MSE,k}k^2|$, as a function of L, and we use the same argument as in the I(1) case to infer that $(\hat{\boldsymbol{\gamma}}_{MSE,L} - \mathbf{b}_{\tilde{x}})'\boldsymbol{\Sigma}^{2\,\prime}$ is absolutely summable. Then, the finite MA-approximation applies if the MA-weights ξ_k decay sufficiently fast for given L. The two equivalent SSA optimization criteria are then obtained from Equation (34), substituting $\boldsymbol{\Sigma}^{2\,\prime}$ for $\boldsymbol{\Sigma}'$

$$\left.\begin{array}{c}\max_{\mathbf{b}_\epsilon}\mathbf{b}_\epsilon'\boldsymbol{\Sigma}^{2\,\prime}\boldsymbol{\gamma}_{\tilde{\Xi}\delta} \\ \mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon = \rho_1\mathbf{b}_\epsilon'\mathbf{b}_\epsilon \\ \mathbf{b}_\epsilon'\boldsymbol{\Sigma}^{2\,\prime}\boldsymbol{\Sigma}^2\mathbf{b}_\epsilon = l\end{array}\right\} \quad \text{or} \quad \left.\begin{array}{c}\min_{\mathbf{b}_\epsilon}(\boldsymbol{\gamma}_{\tilde{\Xi}\delta} - \boldsymbol{\Sigma}^2\mathbf{b}_\epsilon)'(\boldsymbol{\gamma}_{\tilde{\Xi}\delta} - \boldsymbol{\Sigma}^2\mathbf{b}_\epsilon) \\ \mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon = \rho_1\mathbf{b}_\epsilon'\mathbf{b}_\epsilon\end{array}\right\},$$

where $\boldsymbol{\gamma}_{\tilde{\Xi}\delta} = \boldsymbol{\Sigma}^2\boldsymbol{\Xi}\hat{\gamma}_{MSE,L}$. The HT constraint controls the rate of zero-crossings of the stationary $y_t - 2y_{t-1} + y_{t-2}$. From the dual result of Corollary (6) we then infer that $y_t - 2y_{t-1} + y_{t-2}$ has the

[11] We can slightly alter the first two filter coefficients to match the constraints, noting that these adjustments vanish as $L \to \infty$.

least number of zero crossings subject to a given tracking accuracy of the target by the predictor and hence $y_t - y_{t-1}$ is 'most monotonic' so that y_t has the least number of inflection points, i.e., y_t is 'smoothest possible'. In a final step, we must impose level and slope constraints to $\mathbf{b}_{\tilde{x}}$: solving for $b_{\tilde{x}0}, b_{\tilde{x}1}$ then leads to $\mathbf{b}_{\tilde{x}} = (\Gamma(0) - \dot{\Gamma}(0))\mathbf{e}_1 + \dot{\Gamma}(0)\mathbf{e}_2 + \mathbf{B}\tilde{\mathbf{b}}$, where $\tilde{\mathbf{b}}$ is a vector of length $L-2$, $\mathbf{B} = \begin{pmatrix} 1 & 2 & 3 & ... & L-2 \\ -2 & -3 & -4 & ... & -(L-1) \\ 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & & & & \\ 0 & 0 & 0 & ... & 1 \end{pmatrix}$ is an $L \cdot (L-2)$ dimensional matrix and $\mathbf{e}_1, \mathbf{e}_2$ are the first two unit-vectors of length L. An extension of Equation (35) is then straightforward.

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] Davies, N., Pate, M. B. and Frost, M. G. (1974). Maximum autocorrelations for moving average processes. *Biometrika* **61**, 199-200.

[6] Hamilton, J.D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics* **100**, 831–843.

[7] Harvey, A. (1989). Forecasting, structural time series models and the Kalman filter. *Cambridge: Cambridge University Press.*

[8] Henderson, R. (1924). A New Method of Graduation. *Transactions of the Actuarial Society of America* **25**, 29–40

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

[10] Kedem, B. (1986) Zero-crossings analysis. *Research report AFOSR-TR-86-0413, Univ. of Maryland.*

[11] Kratz, M. (2006) Level crossings and other level functionals of stationary Gaussian processes. *Probability surveys* **Vol. 3**, 230-288.

[12] McElroy, T. (2008) Matrix formulas for nonstationary ARIMA signal extraction. *Econometric Theory* **24**, 1—22.

[13] 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.

[14] McElroy, T. and Wildi , M. (2016) Optimal Real-Time Filters for Linear Prediction Problems. *Journal of Time Series Econometrics* **8**, 155-192.

[15] Rice,S.O. (1944) Mathematical analysis of random noise. *I. Bell. Syst. Tech. J* **23**, 282-332.

[16] Whittaker, E.T. (1922). On a New Method of Graduation. *Proceedings of the Edinburgh Mathematical Society* , **41** , 63 - 75.

[17] Wildi, M. (2024) Business-Cycle Analysis and Zero-Crossings of Time Series: a Generalized Forecast Approach. Submitted for publication to the Journal of Business Cycle Analysis.