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

January 7, 2024

Abstract

We propose a generic forecast approach which merges various facets of the prediction problem in terms of mean-square error, sign-accuracy, zero-crossings and smoothness. The latter is formalized by a novel holding-time constraint which conditions the frequency of sign-changes by the predictor. Zero-crossings or sign-changes of a time series can be influential in the decision-making, for e.g. economic actors, by marking transitions between up- and down-turns, expansions and recessions, bull and bear markets, and our forecast approach contributes to such a design of the predictor. The solution to this problem has a simple structural form which allows for a comprehensive analysis of regular, boundary and singular cases as well as for a straightforward derivation of the predictor's distribution. Despite its actual simplicity, the predictor is feature-rich, as illustrated by a series of reproducible examples.

This is a working paper version of SSA univariate: sections 1 and 2 are old; full technical results/proofs are provided in section 3.

1 Introduction

Time series forecasting aims at a coherent analysis of the main systematic dynamics of a phenomenon in view of synthesizing information about future events. Typically, the forecast process is structured by a formal optimality concept, whereby a particular forecast-error measure, such as e.g. the mean-square error (MSE), is minimized. We here argue that multiple and various characteristics of a predictor might draw attention such as the smoothing capability, i.e. the extent by which undesirable 'noisy' components of a time series are suppressed, or timeliness, as measured by relative lead or lag properties of a predictor, or sign accuracy and zero-crossings, as measured by the ability to predict the correct sign of a target. For that purpose, we here propose a generic forecast approach, referred to as simple sign-accuracy (SSA), by merging sign-accuracy and MSE performances subject to a holding-time constraint which determines the expected number of zero-crossings of the predictor in a time interval of given length. Zero-crossings (of the growth-rate) of a time series are influential in the decision-making, for e.g. economic actors, by marking transitions between up- and down-turns, expansions and recessions, bull and bear markets, and our forecast approach contributes to such a design of the predictor. McElroy and Wildi (2019) propose an alternative methodological framework for addressing specific facets of the forecast problem but their approach does not account for zero-crossings explicitly which may be viewed as a shortcoming in some applications. Wildi (2023) proposes an application of SSA to a (real-time) business-cycle analysis, but the chosen treatment remains largely informal. We here fill this gap by providing a complete formal treatment, including regular, singular and boundary cases, a discussion of numerical aspects as well as a derivation of the sample distribution of the predictor together with a comprehensive illustration of technical features and peculiarities of the approach.

The analysis of zero-crossings has been pioneered by Rice (1944). Kedem (1986) and Barnett (1996) extend the concept to exploratory and inferential statistics and a theoretical overview is provided by Kratz (2006). Application fields are various in electronics and image processing, process discrimination, pattern detection in speech, music, or radar screening. However, in contrast to the analysis of current or past events, we here emphasize foremost a prospective prediction perspective.

The optimization criterion is derived in section 2 with a discussion of robustness and extensions of the basic methodological framework (BMF); solutions of the criterion are proposed in section 3 with a discussion of boundary and singular cases, numerical aspects as well as the sample distribution of the estimate; section 4 illustrates applications such as ordinary forecasting, elements of signal extraction and filtering, resilience against departures from the BMF, smoothing and 'unsmoothing', a smoothness-timeliness dilemma, multiplicity and uniqueness features as well as a fully fleshed-out singular case. All empirical examples are reproducible in an open source R-package (include link to Github). Finally, section 5 concludes by summarizing our main findings.

2 Simple Sign-Accuracy (SSA-) Criterion

2.1 Introduction

We propose a simple BMF for presentation of our main results. Specifically, let $\epsilon_t, t \in \mathbb{Z}$, be Gaussian standard white noise[1] and let $\gamma_k \in \mathbb{R}$ for $k \in \mathbb{Z}$ be a square summable sequence $\sum_{k=-\infty}^{\infty} \gamma_k^2 < \infty$. Then $z_t = \sum_{k=-\infty}^{\infty} \gamma_k \epsilon_{t-k}$ is a stationary Gaussian zero-mean process with variance $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We consider estimation of $z_{t+\delta}$, $\delta \in \mathbb{Z}$, referred to as the *target*, based on the predictor $y_t := \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$, where $\mathbf{b} := (b_k)_{0 \le k \le L-1}$ are the coefficients of a finite-length one-sided causal filter: in applications we can set $L = T$, where T is the sample-size, see below for further discussion about the selection of L. This problem is commonly referred to as fore-, now- or backcast, depending on $\delta > 0$, $\delta = 0$ or $\delta < 0$. For illustration, let $z_t = \epsilon_t + \epsilon_{t-1} + \epsilon_{t-2}$ be an MA(2)-process and consider a one-step ahead forecast $\hat{z}_{t+1} = y_t = \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$ of z_{t+1}, see section 4.1 for background. In this case, $\gamma_k = \begin{cases} 1 & 0 \le k \le 2 \\ 0 & \text{otherwise} \end{cases}$, $\delta = 1$, $y_t = \epsilon_t + \epsilon_{t-1}$, $b_0 = b_1 = 1$ and $L = 2$. While classic MSE-predictors can be obtained straightforwardly, at least in this particular case, we here consider a generalization of the MSE-paradigm emphasizing alternative forecast priorities, see section 2.2 for exposition. Departures from the Gaussian assumption will be discussed in sections 2.2 and 4.6 and an extension to autocorrelated stationary x_t such that

$$z_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k} \ , \ x_t = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k} \tag{1}$$

is proposed in section 2.3 with applications in section 4: the coefficients ξ_k in 1 are related to the (purely non-deterministic) Wold-decomposition of x_t. In this more general framework, the above MA(2)-forecast problem can be expressed alternatively by stipulating $\gamma_k = \begin{cases} 1 \ , & k = 0 \\ 0 \ , & \text{otherwise} \end{cases}$ and $\xi_k = \begin{cases} 1 \ , & k = 0, 1, 2 \\ 0 \ , & \text{otherwise} \end{cases}$. The BFM corresponds to $\xi_k = \begin{cases} 1 \ , & k = 0 \\ 0 \ , & \text{otherwise} \end{cases}$ and its extension by 1 addresses more specifically signal extraction, whereby a particular acausal filter $\boldsymbol{\gamma} = (\gamma_k)_{|k|<\infty}$ is applied to an autocorrelated time series x_t in order to extract pertinent 'components'[2]. For illustration, Wildi (2023) considers application of a bi-infinite symmetric HP-filter γ_k^{HP}, see Hodrick and Prescott (1997), to log-returns $x_t = \log(u_t) - \log(u_{t-1})$ of a monthly macro-indicator u_t

[1] Since zero-crossings of zero-mean stationary processes are insensitive to the scaling, our approach is insensitive to σ^2: for simplicity, we will assume $\sigma^2 = 1$ if not stated otherwise.

[2] In principle, 1 could be extended to non-stationary integrated x_t but we shall mainly consider 'stationary' data, for reasons to be explained below. In any case, our procedure could be applied to suitably differenced data and then transformed back via integration to original levels.

and the 'predictor' $y_t = \sum_{k=0}^{L-1} b_k x_{t-k}$ of $z_{t+\delta}$ is designed for a nowcast ($\delta = 0$) of the cycle $z_t = \sum_{|k|<\infty} \gamma_k^{HP} x_{t-k}$ at the current sample end, $t = T$, see also section 4 for worked-out examples. To conclude, let us clarify that x_t is the observed data, z_t is a generally unobserved component or filtered version of x_t; moreover, we assume γ_k and ξ_j in 1 to be known, either 'a priori', for example by selecting a known HP-filter, or because they have been computed in advance, by classic time series analysis techniques. Given 1, we then here propose a generalized prediction criterion for deriving $\mathbf{b}$ of y_t which merges alternative forecast priorities, including sign-accuracy, mean-square performances and 'smoothness', to be defined below. In any case, the BFM chiefly intends to clarify exposition and to simplify notation in view of highlighting relevant facets of the gereralized prediction problem in a decluttered formal context: extensions will be addressed in sections 2.2, 2.3 and 4.

2.2 Sign-Accuracy, MSE and Holding-Time

To begin, we assume pertinence of the basic methodological framework (BMF) and look for an estimate y_t of $z_{t+\delta}$, given data $\epsilon_t, \epsilon_{t-1}, ...$ and filter-coefficients $\gamma_k, |k| < \infty$, such that the probability $\mathrm{P}\Big(\mathrm{sign}(z_{t+\delta}) = \mathrm{sign}(y_t)\Big)$ is maximized as a function of $\mathbf{b}$: we refer to this criterion in terms of *sign accuracy* (SA).

Proposition 1 *Under the BMF the sign accuracy criterion can be stated as*

$$\max_{\mathbf{b}} \rho(y, z, \delta) \tag{2}$$

where

$$\rho(y, z, \delta) = \frac{\sum_{k=0}^{L-1} \gamma_{k+\delta} b_k}{\sqrt{\sum_{k=-\infty}^{\infty} \gamma_k^2} \sqrt{\sum_{k=0}^{L-1} b_k^2}}$$

is the correlation between y_t and $z_{t+\delta}$.

In the stipulated case of Gaussian random variables a proof follows readily from the identity $P\Big(\mathrm{sign}(z_{t+\delta}) = \mathrm{sign}(y_t)\Big) = 0.5 + \frac{\arcsin(\rho(y,z,\delta))}{\pi}$, relying on strict monotonicity of the non-linear transformation. We then infer that SA and MSE are equivalent criteria, at least down to an arbitrary scaling of y_t and conditional on the Gaussian assumption.

Remarks

We here discard the scaling parameter from further consideration since our approach emphasizes signs, smoothness and timeliness aspects as alternative priorities. In this perspective, predictors that differ by an arbitrary (positive) normalization constant are felt equivalent. Also, computation of the optimal MSE-scaling would be a simple exercise, if needed. Note that classification methods such as e.g. logit models are less suitable for the purpose at hand because fitting the signs $\mathrm{sign}(z_{t+\delta}) = \pm 1$, instead of the actual observations $z_{t+\delta}$, would result in a loss of efficiency under the premises of the BMF.

Consider now the expected duration between consecutive zero-crossings or sign-changes of the predictor y_t, which will be referred to as *holding-time*.

Proposition 2 *Under the BMF the holding-time $ht(y|\mathbf{b})$ of y_t is*

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

where $\rho(y, y, 1) = \frac{\sum_{i=1}^{L-1} b_i b_{i-1}}{\sum_{i=0}^{L-1} b_i^2}$ is the lag-one autocorrelation of y_t.

A proof is provided by Kedem (1986). We can now formalize the concept of 'smoothness' of a predictor y_t by constraining $\mathbf{b}$ such that

$$ht(y|\mathbf{b}) = ht_1 \tag{4}$$

or, equivalently,

$$\rho(y, y, 1) = \rho_1 \tag{5}$$

where ht_1 or ρ_1, linked through 3, are proper hyper-parameters of our design. In the following, we refer to the 'holding-time' either in terms of $ht(y|\mathbf{b})$ or $\rho(y, y, 1)$, clarifying our intent in case of ambiguity. We here argue that the hyper-parameter ht_1 is interpretable and can be set a priori, at the onset of an analysis, according to structural elements of a prediction problem. As an example, Wildi (2023) illustrates the proceeding in a business-cycle application, where ht_1 matches the length of historical recession episodes. Also, the holding-time could be selected in view of taming the number of unsystematic or noisy crossings (false alarms). Furthermore, if costly strategy-adjustments or behavioral changes take place at zero-crossings, for example in an algorithmic trading framework or for macro-economic policy setting, then cumulated costs of all adjustments, being inversely proportional to ht_1, could be accounted for by a suitable determination of the hyperparameter. Finally, ht_1 could be set according to short-, mid- or long-term i.e tactic, strategic or fundamental outlook perspectives. Concerning the proper selection of the holding-time, the following proposition sets limits for admissible constraints in the basic framework.

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

We refer to N. Davies, M. B. Pate and M. G. Frost (1974) for a proof, see also proposition 5 further down. Consider now the sign accuracy criterion 2 endowed with the holding-time constraint 5:

$$\left.\begin{array}{l} \max_{\mathbf{b}} \frac{\sum_{k=0}^{L-1} \gamma_{k+\delta} b_k}{\sqrt{\sum_{k=-\infty}^{\infty} \gamma_k^2} \sqrt{\sum_{k=0}^{L-1} b_k^2}} \\ \frac{\sum_{k=1}^{L-1} b_{k-1} b_k}{\sum_{k=0}^{L-1} b_k^2} = \rho_1 \end{array}\right\} \tag{6}$$

This optimization problem is called *simple sign-accuracy* or SSA-criterion: simplicity here refers to the elementary structure of the predictor, as derived in theorem 1, as well as to the scope of the criterion which does not yet allow for a formal treatment of timeliness or lead/lag issues, see section 4.7 for an informal treatment and Wildi (2023) for additional illustration. We allude to solutions of this criterion by the acronym SSA or SSA(ht_1, δ) or SSA(ρ_1, δ) to stress the dependence of the predictor on the pair of hyper-parameters, see section 4.7 for reference. The SSA-criterion merges MSE, sign accuracy and smoothing requirements in a flexible and consistent way. Departures from the Gaussian assumption can be accommodated in the sense that y_t or z_t can be 'nearly Gaussian' even if $x_t = \epsilon_t$ is not, due to the central limit theorem, see Wildi (2023) for an application to financial data (equity index) and section 4.6. Finally, the criterion remains appealing outside of a strict holding-time or zero-crossing perspective by complementing the classic predictor with a generic smoothing constraint addressing the lag-one autocorrelation function (acf).

2.3 Extension to Stationary Processes

Let

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

be stationary Gaussian processes and designate by ξ_i the weights of the (purely non-deterministic) Wold-decomposition of x_t. Then target and predictor can be formally re-written as

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

where $(\gamma \cdot \xi)_k = \sum_{m \leq k} \xi_{k-m} \gamma_m$ and $(b \cdot \xi)_j = \sum_{n=0}^{\min(L-1,j)} \xi_{j-n} b_n$ are convolutions of the sequences γ_k and b_j with the Wold-decomposition ξ_i of x_t. The SSA-criterion then becomes

$$\begin{aligned} \max_{(\mathbf{b} \cdot \boldsymbol{\xi})} & \frac{\sum_{k \geq 0} (\gamma \cdot \xi)_{k+\delta} (b \cdot \xi)_k}{\sqrt{\sum_{|k|<\infty} (\gamma \cdot \xi)_k^2} \sqrt{\sum_{j \geq 0} (b \cdot \xi)_j^2}} \\ & \frac{\sum_{j \geq 1} (b \cdot \xi)_{j-1} (b \cdot \xi)_j}{\sum_{j \geq 0} (b \cdot \xi)_j^2} = \rho_1 \end{aligned} \tag{7}$$

which can be solved for $(b \cdot \xi)_j, j = 0, 1, ...$, see theorem 1. The sought-after filter coefficients b_k can then be obtained from $(b \cdot \xi)_j$ by inversion or deconvolution

$$b_k = \begin{cases} (b \cdot \xi)_0 / \xi_0 & , k = 0 \\ \frac{(b \cdot \xi)_k - \sum_{j=0}^{k-1} \xi_{k-j} b_j}{\xi_0} & k > 0 \end{cases} \tag{8}$$

assuming $\xi_0 \neq 0$. Otherwise, if $\xi_k = 0, k = 0, ..., k_0 - 1$ and $\xi_{k_0} \neq 0$, then the inversion is initialized with $b_0 = (b \cdot \xi)_{k_0} / \xi_{k_0}$. Note that non-stationary integrated processes could be addressed in a similar vein, assuming some initialization settings, see e.g. McElroy and Wildi (2020). However, since the concept of a holding-time, i.e. the expected duration between consecutive zero-crossings, would generally not be properly defined anymore, we henceforth assume non-stationary trending data to be suitably transformed or differenced. Also, we refer to standard results in textbooks for a derivation of ξ_k or ϵ_t based on a finite sample of observations $x_1, ..., x_T$, see e.g. Brockwell and Davis (1993): worked-out examples are provided in sections 4.2, 4.3 and 4.4. Finally, for notational convenience we henceforth rely on the BMF, acknowledging that straightforward modifications would apply in the case of autocorrelation.

???? Generalize to ARMA by relying on MA-inversion in my TS-lecture: use first order terms only.

Let $x_t = a_1 x_{t-1} + \epsilon_t$ denote an AR(1)-process and let $\Delta := a_1 - E[\hat{a}_1]$ designate the (finite sample) estimation bias of the classic MSE-estimate $\hat{a}_1$ of a_1. We first analyze the effect of a small bias Δ on the lag-one acf of the SSA-filter b_k applied to x_t. First note that

$$y_t = \sum_{k=0}^{L-1} b_k x_{t-k} = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{k} b_j a_1^{k-j} \right) \epsilon_{t-k}$$

where we assume that $b_j = 0$ for $j > L$. We then infer

$$\begin{aligned}\rho(y,y,1) &= \frac{\sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} b_j(a_1+\Delta)^{k-j}\right)\left(\sum_{j=0}^{k+1} b_j(a_1+\Delta)^{k+1-j}\right)}{\sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} b_j(a_1+\Delta)^{k-j}\right)^2} \\ &\approx \frac{b_0(a_1+\Delta)}{\sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} b_j(a_1+\Delta)^{k-j}\right)^2}\end{aligned}$$

so that $\partial\rho(y,y,1) \approx$???. We know derive the effect of the bias on the holding-time. From

$$ht = \frac{\pi}{\arccos(\rho(y,y,1))}$$

we infer

$$\frac{\partial ht}{\partial \Delta} = -\frac{\pi}{\arccos(\rho(y,y,1))^2}\frac{-1}{\sqrt{1-\rho(y,y,1)^2}}\partial\rho(y,y,1)$$

3 Solution of the SSA-Criterion

The following proposition re-formulates the target specification in terms of the MSE-predictor.

Proposition 4 *Under the BMF let* $\hat{z}_{t,\delta} = \sum_{k=0}^{L-1}\gamma_{k+\delta}\epsilon_{t-k} = \boldsymbol{\gamma}_\delta'\boldsymbol{\epsilon}_t$ *designate the classic MSE-predictor of* $z_{t+\delta}$*. Then the original target* $z_{t+\delta}$ *can be replaced by* $\hat{z}_{t,\delta}$ *in the SSA-criterion.*

Proof

A proof follows from

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

where $\cdot|\rho_1$ denotes conditioning, subject to the holding-time constraint, and $\text{Arg}(\cdot)$ means the solution or argument of the optimization.

The proposition suggests that the SSA-predictor y_t should 'fit' the MSE-predictor $\hat{z}_{t,\delta}$ while complying with the holding-time constraint: if $\hat{z}_{t,\delta}$ matches the constraint then SSA and MSE coincide, up to arbitrary scaling, and the constraint could be dropped (so-called 'degenerate' case, see second regularity assumption of theorem 1 below). Therefore, we henceforth refer to $\hat{z}_{t,\delta}$ (or $\boldsymbol{\gamma}_\delta$) as an equivalent target specification. In the following we assume that $\boldsymbol{\gamma}_\delta \neq 0$, see also the first regularity assumption of theorem 1 below. Otherwise, if $\boldsymbol{\gamma}_\delta = 0$, then $\rho(y,z,\delta) = 0$ for all $\mathbf{b}$ i.e. the objective function of the SSA-criterion would vanish irrespective of y_t and therefore the SSA-solution would not be properly identified anymore: all $\mathbf{b} \in \mathbb{R}^L$ would be equally valid or invalid 'solutions'.

Consider next the so-called autocovariance-generating matrix

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

of dimension $L * L$ so that $\rho(y, y, 1) = \dfrac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}$. The following proposition relates stationary points of the lag-one autocorrelation $\rho(y, y, 1)$ to eigenvectors and eigenvalues of $\mathbf{M}$.

Proposition 5 *Under the BMF, the vector* $\mathbf{b} := (b_0, ..., b_{L-1})' \neq 0$ *is a stationary point of the lag-one autocorrelation* $\rho(y, y, 1) = \dfrac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}$ *if and only if* $\mathbf{b}$ *is an eigenvector of the autocovariance-generating matrix with corresponding eigenvalue* $\rho(y, y, 1)$. *The extremal values* $\rho_{min}(L)$ *and* $\rho_{max}(L)$ *defined in proposition 3 correspond to* $\min_i \lambda_i$ *and* $\max_i \lambda_i$ *where* λ_i, $i = 1, ...L$ *are the eigenvalues of* $\mathbf{M}$.

Proof

Assume, for simplicity, that $\mathbf{b} \neq \mathbf{0}$ is defined on the unit-sphere so that

$$\begin{aligned} \mathbf{b}'\mathbf{b} &= 1 \\ \rho(y, y, 1) &= \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \mathbf{b}'\mathbf{M}\mathbf{b} \end{aligned}$$

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.

$$\mathbf{M}\mathbf{b} = \lambda\mathbf{b}$$

We deduce that $\mathbf{b}$ is a stationary point if and only if it is an eigenvector of $\mathbf{M}$. Then

$$\rho(y, y, 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \lambda_i \frac{\mathbf{b}'\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \lambda_i$$

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 i.e. $\rho_{min}(L) = \min_i \lambda_i$ and $\rho_{max}(L) = \max_i \lambda_i$.

By abuse of terminology we now identify filter coefficients and corresponding filter outputs (targets or predictors) so that e.g. y_t and $\mathbf{b}$ will both be referred to as predictor or estimate (and similarly for the target(s)). Let then $\lambda_j, \mathbf{v}_j$ denote the pairings of eigenvalues and eigenvectors of $\mathbf{M}$, ordered according to the increasing size of $\lambda_j = -\cos(\omega_j)$, where $\omega_j = j\pi/(L + 1)$ are the discrete Fourier frequencies see e.g. Anderson (1975), and let $\mathbf{V}$ designate the orthonormal basis of $\mathbb{R}^L$ based on the (Fourier) column-vectors $\mathbf{v}_j$, $j = 1, ..., L$. We then consider the spectral decomposition of the target $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$

$$\boldsymbol{\gamma}_\delta = \sum_{i=n}^{m} w_i \mathbf{v}_i = \mathbf{V}\mathbf{w} \tag{9}$$

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 $\boldsymbol{\gamma}_\delta$ is called *band-limited*. Also, we refer to $\boldsymbol{\gamma}_\delta$ as having *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 $NZ = \{1, 2, ..., L\}$ or $NZ \subset \{1, 2, ..., L\}$ depending on $\boldsymbol{\gamma}_\delta$ having complete or incomplete spectral support. The following theorem derives a parametric functional form of the SSA solution under various assumptions about the problem specification.

Theorem 1 *Consider the SSA optimization problem 6 under the BMF and consider the following set of regularity assumptions:*

1. $\boldsymbol{\gamma}_\delta \neq 0$ *(identifiability) and* $L \geq 3$ *(smoothing).*
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 of the holding-time constraint).*

4. *The MSE-estimate* $\boldsymbol{\gamma}_\delta$ *has complete spectral support (completeness).*

Then

1. *If the third regularity assumption is violated (admissibility) and if* $|\rho_1| > \rho_{max}(L)$*, then the problem cannot be solved unless the filter-length* L *is increased such that* $|\rho_1| \leq \rho_{max}(L)$*. On the other hand, if* $\rho_1 = \lambda_1 = -\rho_{max}(L)$ *or* $\rho_1 = \lambda_L = \rho_{max}(L)$ *(limiting cases), then* $sign(w_1)\mathbf{v}_1$ *or* $sign(w_L)\mathbf{v}_L$ *are the corresponding solutions of the SSA-criterion (up to arbitrary scaling), where* w_i *are the spectral weights in 9 and where it is assumed that* $w_1 \neq 0$*, if* $\rho_1 = \lambda_1$*, or* $w_L \neq 0$*, if* $\rho_1 = \lambda_L$*.*

2. *If all regularity assumptions hold, then the SSA-estimate* $\mathbf{b}$ *has the parametric functional form*

$$\mathbf{b} = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta = D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \tag{10}$$

where $D \neq 0$*,* $\nu \in \mathbb{R} - \{2\lambda_i | i = 1, ..., L\}$*, and* $\boldsymbol{\nu} := 2\mathbf{M} - \nu\mathbf{I}$ *is an invertible* $L * L$ *matrix. Although* b_{-1}, b_L *do not explicitly appear in* $\mathbf{b}$ *it is at least implicitly assumed that* $b_{-1} = b_L = 0$ *(implicit boundary constraints). Furthermore,* $\mathbf{b}$ *is uniquely determined by the scalar* ν*, down to the arbitrary scaling term* D*, whereby the sign of* D *is determined by requiring a positive (SSA-) criterion-value.*

3. *If all regularity assumptions hold, then the lag-one autocorrelation of* $\mathbf{b}$ *in 10 is*

$$\rho(\nu) := \rho(y(\nu), y(\nu), 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\sum_{i=1}^{L}\lambda_i w_i^2 \frac{1}{(2\lambda_i - \nu)^2}}{\sum_{i=1}^{L} w_i^2 \frac{1}{(2\lambda_i - \nu)^2}} \tag{11}$$

and $\nu = \nu(\rho_1)$ *can always be selected such that the SSA-solution* $\mathbf{b} = \mathbf{b}(\nu(\rho_1))$ *in 10 complies with the holding-time constraint.*

4. *For* $\nu \in \{x | |x| > 2\rho_{max}(L)\}$ *the lag-one ACF* $\rho(\nu)$ *defined in 11 is a bijective function of* ν *and the parameter* ν *in 10 is determined uniquely by the holding-time constraint* $\rho(\nu) = \rho_1$*. The function* $\rho(\nu)$ *is differentiable in* ν *and* $\partial\rho(\nu)/\partial\nu < 0$ *in* $\nu \in \{x | |x| > 2\rho_{max}(L)\}$*. Finally,* $max_{\nu < -2\rho_{max}(L)}\rho(\nu) = min_{\nu > 2\rho_{max}(L)}\rho(\nu) = \rho_{MSE}$*, the lag-one ACF of* $\boldsymbol{\gamma}_\delta$*.*

5. *For* $\nu \in \{x | x > 2\rho_{max}(L)\}$ *or for* $\nu \in \{x | x < -2\rho_{max}(L)\}$ *the target correlation* $\rho(y(\nu), z, \delta)$*, where* $y_t(\nu)$ *designates the output of the filter* $\mathbf{b} = \mathbf{b}(\nu)$ *based on 10, is a strictly monotonic function in* ν*. Moreover,* $\rho(y(\nu), z, \delta)$ *and* $\rho(\nu)$ *are linked by*

$$-sign(\nu)\frac{\partial\rho(y(\nu), z, \delta)}{\partial\nu} = \frac{1}{\left(\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1}\ '\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta\right)^{3/2}\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\frac{\partial\rho(\nu)}{\partial\nu} < 0 \tag{12}$$

Proof

The SSA-problem 6 can be rewritten as

$$\begin{aligned} \max_{\mathbf{b}}\ & \boldsymbol{\gamma}_\delta'\mathbf{b} \\ & \mathbf{b}'\mathbf{b} = 1 \\ & \mathbf{b}'\mathbf{M}\mathbf{b} = \rho_1 \end{aligned} \tag{13}$$

where $\mathbf{b}'\mathbf{b} = 1$ is an arbitrary scaling rule. Consider the spectral decomposition

$$\mathbf{b} := \sum_{i=1}^{L}\alpha_i\mathbf{v}_i \tag{14}$$

of $\mathbf{b}$. Since $\mathbf{v}_i$ is an orthonormal basis, the length-constraint $\mathbf{b}'\mathbf{b} = 1$ implies $\sum_{i=1}^{L} \alpha_i^2 = 1$ (unit-sphere constraint); moreover, from the holding-time constraint and from orthogonality of $\mathbf{v}_i$ we infer

$$\rho_1 = \mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{i=1}^{L} \alpha_i^2 \lambda_i$$

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 $\lambda_{j_0} \neq 0$[3]. The SSA-problem can be solved if the hyperbola, defined by the holding-time constraint, intersects the unit-sphere. For this purpose we plug the former equation into the latter:

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

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

$$\alpha_{i_0} = \pm \sqrt{\frac{\lambda_{j_0} - \rho_1}{\lambda_{j_0} - \lambda_{i_0}} - \sum_{k \neq i_0, k \neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}}} \tag{15}$$

Under the limiting cases posited in assertion 1 $\rho_1 = \lambda_{i_0}$ with either $i_0 = 1$, i.e. $\rho_1 = -\rho_{max}(L)$, or $i_0 = L$, i.e. $\rho_1 = \rho_{max}(L)$. Let then $i_0 = 1$ so that 15 becomes

$$\alpha_1 = \pm \sqrt{1 - \sum_{k \neq 1, k \neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_1}} \tag{16}$$

Assume also $j_0 = 2$ (a similar proof can be derived for arbitrary $j_0 \geq 2$, see footnote 4) so that $\lambda_2 - \lambda_k < 0$ in the nominator and $\lambda_2 - \lambda_1 > 0$ in the denominator of 16. Therefore, the term under the square-root is larger than one if $\alpha_k \neq 0$ for some $k > 2$ which would imply $|\alpha_1| > 1$ thus contradicting the unit-sphere constraint[4]. We then deduce $\alpha_k = 0$ for $k > 2$ so that $\alpha_1 = \pm 1$ and $\alpha_2 = 0$ and therefore $\pm\mathbf{v}_1$ are the only admissible potential solutions of the SSA-problem: the contacts of unit-sphere and hyperbola are tangential at the vertices $\pm\mathbf{v}_1$. Since $w_1 \neq 0$ by assumption, the solution must be $\mathbf{b} := \text{sign}(w_1)\mathbf{v}_1$ because it maximizes the criterion value $\boldsymbol{\gamma}_\delta'\mathbf{b} = \text{sign}(w_1)w_1 > 0$. If $w_1 = 0$ then the problem is ill-conditioned in the sense that the only possible solutions $\pm\mathbf{v}_1$ do not correlate with the target $z_{t+\delta}$ anymore. Note that similar reasoning applies if $i_0 = L$, setting $j_0 = L - 1$ in 15 and assuming $w_L \neq 0$.

To show the second assertion we now assume that all regularity assumptions hold and we define the Lagrangian function

$$L := \boldsymbol{\gamma}_\delta'\mathbf{b} - \lambda_1(\mathbf{b}'\mathbf{b} - 1) - \lambda_2(\mathbf{b}'\mathbf{M}\mathbf{b} - \rho_1) \tag{17}$$

By assumption $L \geq 3$ (smoothing) so that $\mathbf{b}$ is defined on a $L - 2 \geq 1$ dimensional intersection of unit-sphere and holding-time constraints, as specified by 15. Since this intersection is free of boundary points, the solution $\mathbf{b}$ of the SSA-problem must conform to the stationary Lagrangian or vanishing gradient equations

$$\boldsymbol{\gamma}_\delta = \lambda_1 2\mathbf{b} + \lambda_2(\mathbf{M} + \mathbf{M}')\mathbf{b} = \lambda_1 2\mathbf{b} + \lambda_2 2\mathbf{M}\mathbf{b} \tag{18}$$

[3] If L is an even integer, then $\lambda_i \neq 0$ for all i, $1 \leq i \leq L$. Otherwise, $\lambda_{i_0} = 0$ for $i_0 = 1 + (L-1)/2$.

[4] Similar contradictions could be derived for any $j_0 > 1$ since $\left| \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}} \right| < 1$ if $i_0 = 1$ so that if any $\alpha_k \neq 0$, for $k \neq 1, j_0$, then equation 15 would conflict with the unit-sphere constraint.

Note that the second regularity assumption (non-degenerate case) implies that the holding-time constraint 13 is 'active' i.e. $\lambda_2 \neq 0$. Dividing by λ_2 then leads to

$$\begin{aligned} D\boldsymbol{\gamma}_\delta &= \boldsymbol{\nu}\mathbf{b} && (19) \\ \boldsymbol{\nu} &:= (2\mathbf{M} - \nu\mathbf{I}) && (20) \end{aligned}$$

where $D = 1/\lambda_2$ and $\nu = -2\frac{\lambda_1}{\lambda_2}$. By orthonormality of $\mathbf{v}_i$ the objective function is

$$\boldsymbol{\gamma}_\delta'\mathbf{b} = \sum_{i=1}^{L} \alpha_i w_i$$

where we rely on the spectral decomposition 14 of $\mathbf{b}$. By assumption $L \geq 3$ (smoothing) so that $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_L)'$ is defined on a $L-2 \geq 1$ dimensional intersection of unit-sphere and holding-time constraints. We then infer that the objective function is not overruled by the constraint i.e. $|\lambda_2| < \infty$ so that $D \neq 0$ in 19, as claimed. Furthermore, equation 19 can be written as

$$\begin{aligned} b_{k+1} - \nu b_k + b_{k-1} &= D\gamma_{k+\delta}\ ,\ 1 \leq k \leq L-2 && (21) \\ b_1 - \nu b_0 &= D\gamma_\delta\ ,\ k = 0 \\ -\nu b_{L-1} + b_{L-2} &= D\gamma_{L-1+\delta}\ ,\ k = L-1 \end{aligned}$$

for $k = 0, ..., L-1$ so that $b_{-1} = b_L = 0$ are implicitly assumed for the natural extension $(b_{-1}, \mathbf{b}, b_L)'$ of the time-invariant linear filter. The eigenvalues of $\boldsymbol{\nu}$ are $2\lambda_i - \nu$ with corresponding eigenvectors $\mathbf{v}_i$. We note that if $\mathbf{b}$ is the solution of the SSA-problem, then $\nu/2$ cannot be an eigenvalue of $\mathbf{M}$ since otherwise $\boldsymbol{\nu}$ in 19 would map one of the eigenvectors in the spectral decomposition of $\mathbf{b}$ to zero which would contradict the last regularity assumption (completeness: see corollary 1 for a corresponding extension) since $D \neq 0$. Therefore we can assume that $\boldsymbol{\nu}^{-1}$ exists and

$$\boldsymbol{\nu}^{-1} = \mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'$$

where the diagonal matrix $\mathbf{D}_\nu^{-1}$ has entries $\frac{1}{2\lambda_i - \nu}$. We can then solve 19 for $\mathbf{b}$ and obtain

$$\begin{aligned} \mathbf{b} &= D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta && (22) \\ &= D\mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'\mathbf{V}\mathbf{w} \\ &= D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i && (23) \end{aligned}$$

where we inserted 9. Since $\boldsymbol{\nu}$ has full rank, the solution of the SSA-problem is uniquely determined by ν, at least down to arbitrary scaling, hereby completing the proof of assertion 2. Note that $\mathbf{b}$ as given by 23 is a linear combination of (Fourier-) eigenvectors $\mathbf{v}_k$ of $\mathbf{M}$: the $j-$th component of $\mathbf{v}_k$ is $\sin(kj\pi/(L+1))/\sqrt{(\sum_{l=1}^{L}(\sin(kl\pi/(L+1))^2)}$, $j = 1, ...L$. Therefore, the artificially extended boundary-values of the vectors at $j = 0$ and $j = L+1$ vanish thereby confirming the implicit boundary constraints $b_{-1} = b_L = 0$ for $\mathbf{b}$ derived above.
We next proceed to assertion 3 and consider

$$\begin{aligned} \rho(\nu) &= \rho(y(\nu), y(\nu), 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} \\ &= \frac{\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)'\mathbf{M}\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)}{\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)'\left(D\sum_{i=1}^{L}\frac{w_i}{2\lambda_i-\nu}\mathbf{v}_i\right)} \\ &= \frac{\sum_{i=1}^{L}\frac{\lambda_i w_i^2}{(2\lambda_i-\nu)^2}}{\sum_{i=1}^{L}\frac{w_i^2}{(2\lambda_i-\nu)^2}} && (24) \end{aligned}$$

where we inserted 23 and made use of orthonormality $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$. The last expression implies $\lim_{\nu\to 2\lambda_i}\rho(\nu) = \lambda_i$ for all $i = 1, ..., L$ such that $\lambda_i \neq 0$. Since $\lambda_1 = -\rho_{max}(L)$ and $\lambda_L = \rho_{max}(L)$, by proposition 5, we infer that lower and upper boundaries $\pm\rho_{max}(L)$ can be reached by $\rho(\nu)$, asymptotically. 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 holding-time constraint under the posited assumptions. We now proceed to assertion 4 by showing that the parameter ν is determined uniquely by ρ_1 in the holding-time constraint if $|\nu| > 2\rho_{max}(L)$. Note that all eigenvalues $2\lambda_i - \nu$ of $\boldsymbol{\nu}$ must be (strictly) negative, if $\nu > 2\rho_{max}(L)$, or strictly positive, if $\nu < -2\rho_{max}(L)$, so that all eigenvalues of $\boldsymbol{\nu}^{-1}$, being the reciprocals of the former, must be of the same sign, either all positive or all negative. Finally, the eigenvalues of $\boldsymbol{\nu}$ or $\boldsymbol{\nu}^{-1}$ must be pairwise different since the eigenvalues of $\mathbf{M}$ are so. We then obtain

$$\begin{aligned}
\frac{\partial\rho\Big(y(\nu),y(\nu),1\Big)}{\partial\nu} &= \frac{\partial}{\partial\nu}\left(\frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}\right) = \frac{\partial}{\partial\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\ \prime}\mathbf{M}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\ \prime}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}\right) = \frac{\partial}{\partial\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}\right)\\
&= \frac{2\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta\mathbf{b}'\mathbf{b}/D - (2\mathbf{b}'\mathbf{M}\mathbf{b}/D)\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta}{((\mathbf{b}'\mathbf{b})^2/D^2)}\\
&= \frac{2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b}/D^2 - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}/D^2}{\mathbf{b}'\mathbf{b}/D^2}\\
&= 2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b} - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b} \qquad (25)
\end{aligned}$$

where $\boldsymbol{\nu}^{-k} := (\boldsymbol{\nu}^{-1})^k$, $\boldsymbol{\nu}^{-1\ \prime} = \boldsymbol{\nu}^{-1}$ (symmetry); commutativity of the matrix multiplications (used in deriving the third and next-to-last equations) follows from the fact that the matrices are symmetric and simultaneously diagonalizable (same eigenvectors); also we relied on generic matrix differentiation rules in the third equation[5]; finally we relied on $\mathbf{b}'\mathbf{b} = 1$ in the last equation. We can now insert

$$\mathbf{M}\boldsymbol{\nu}^{-1} = \frac{\nu}{2}\boldsymbol{\nu}^{-1} + 0.5\mathbf{I}$$

which is a reformulation of $(2\mathbf{M} - \nu\mathbf{I})\boldsymbol{\nu}^{-1} = \mathbf{I}$ into the first summand in 25 to obtain

$$2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b} = \left(\nu\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b} + \mathbf{b}'\mathbf{b}\right)\mathbf{b}'\mathbf{b}$$

We can now insert this expression into 25 and isolate $\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}$ to obtain

$$\begin{aligned}
\frac{\partial\rho\Big(y(\nu),y(\nu),1\Big)}{\partial\nu} &= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\left(2\mathbf{b}'\mathbf{M}\mathbf{b} - \nu\mathbf{b}'\mathbf{b}\right) + (\mathbf{b}'\mathbf{b})^2\\
&= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\left(2\mathbf{M} - \nu\mathbf{I}\right)\mathbf{b} + (\mathbf{b}'\mathbf{b})^2\\
&= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}\mathbf{b} + (\mathbf{b}'\mathbf{b})^2\\
&= -\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta + (\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta)^2 \qquad (26)\\
&= -\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-3}\mathbf{V}'\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-1}\mathbf{V}'\boldsymbol{\gamma}_\delta + (\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-2}\mathbf{V}'\boldsymbol{\gamma}_\delta)^2\\
&= -\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-3}\tilde{\boldsymbol{\gamma}}_{+\delta}\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-1}\tilde{\boldsymbol{\gamma}}_{+\delta} + (\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-2}\tilde{\boldsymbol{\gamma}}_{+\delta})^2
\end{aligned}$$

where $\boldsymbol{\nu}^{-k} = \mathbf{V}\mathbf{D}^{-k}\mathbf{V}'$ and $\mathbf{D}^{-k}$, $k = 1, 2, 3$, is diagonal with eigenvalues $\lambda_{i\nu}^{-k} := (2\lambda_i - \nu)^{-k}$ being all (strictly) positive, if $\nu < -2\rho_{max}(L)$, or either all (strictly) negative or all (strictly) positive depending on the exponent k being odd or even, if $\nu > 2\rho_{max}(L)$; also, $\tilde{\boldsymbol{\gamma}}_{+\delta} = \mathbf{V}'\boldsymbol{\gamma}_{+\delta} = (w_1, ..., w_L)'$. Therefore

$$\begin{aligned}
\frac{\partial\rho\Big(y(\nu),y(\nu),1\Big)}{\partial\nu} &= -\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-3}\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-1} + \left(\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-2}\right)^2\\
&= -\sum_{i>k} w_i^2 w_k^2\Big(\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-3} + \lambda_{i\nu}^{-3}\lambda_{k\nu}^{-1} - 2\lambda_{i\nu}^{-2}\lambda_{k\nu}^{-2}\Big) \qquad (27)
\end{aligned}$$

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

where the terms in w_j^4 cancel. Consider now

$$\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-3}+\lambda_{i\nu}^{-3}\lambda_{k\nu}^{-1}-2\lambda_{i\nu}^{-2}\lambda_{k\nu}^{-2} \quad = \quad \lambda_{i\nu}^{-1}\lambda_{k\nu}^{-1}\Big(\lambda_{i\nu}^{-2}+\lambda_{k\nu}^{-2}-2\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-1}\Big)=\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-1}\Big(\lambda_{i\nu}^{-1}-\lambda_{k\nu}^{-1}\Big)^2 \;>\; 0$$

where the strict inequality holds because $\lambda_{i\nu}^{-1}=(2\lambda_i-\nu)^{-1}$ are all of the same sign, pairwise different and non-vanishing if $|\nu|>2\rho_{max}(L)$. Since $w_i\neq 0$ (last regularity assumption: completeness) we deduce $w_i^2w_k^2\neq 0$ in 27. Therefore, the latter expression 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 $\nu\in\{x|x<-2\rho_{max}(L)\}$. In order to show bijectivity for $\nu\in\{x||x|>2\rho_{max}(L)\}$ we note that $\lim_{|\nu|\to\infty}\rho(\nu)=\frac{\sum_{i=1}^L\lambda_i w_i^2}{\sum_{i=1}^L w_i^2}$, see 24, and the limiting value corresponds to the lag-one ACF of the MSE predictor $\boldsymbol{\gamma}_\delta$. But $\lim_{\nu\to\infty}\rho(\nu)=\lim_{\nu\to-\infty}\rho(\nu)$ together with $\frac{\partial\rho(\nu)}{\partial\nu}<0$ imply bijectivity for $\nu\in\{x||x|>2\rho_{max}(L)\}$; also, $\max_{\nu<-2\rho_{max}(L)}\rho(\nu)=\min_{\nu>2\rho_{max}(L)}\rho(\nu)=\rho_{MSE}$, as claimed. Finally, we examine assertion 5. First

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

For $\nu<-2\rho_{max}(L)$ the quotient is strictly positive and $\text{sign}(D)>0$; for $\nu>2\rho_{max}(L)$ the quotient is strictly negative and $\text{sign}(D)<0$. Assume now, that $\nu<-2\rho_{max}(L)$ so that

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

The last equality is obtained by recognizing that the expression in curly brackets is identical with 26. The inequality follows from the proof of assertion 4, since the scaling term is strictly positive. If $\nu>2\rho_{max}(L)$ the same proof applies, but with changed sign, $\text{sign}(D)=1$, and accordingly modified strict inequality, at the end.

Remarks

Extensions to autocorrelated x_t are straightforward, see section 2.3 for background and sections 4.2, 4.3 and 4.4 for illustration. Also, Gaussianity is not required in the derivation of the above proof because the SSA-criterion 6 relies solely on correlations. The Gaussian hypothesis is needed when establishing formal links between correlations and sign-accuracy or holding-time concepts but y_t or z_t can be nearly Gaussian even if ϵ_t isn't, see also section 4.6 for illustration. More generally, the proof applies to constraints of the form $\mathbf{b}'\tilde{\mathbf{M}}\mathbf{b}=c\mathbf{b}'\mathbf{b}$ for arbitrary symmetric $\tilde{\mathbf{M}}$, with pairwise different eigenvalues $\tilde{\lambda}_i$ whereby pairwise difference is required for a proof of the last assertion only. Therefore, more general constraints could be considered, involving e.g. a linear combination of acfs at various lags or the entire spectrum of the process instead of the lag-one holding-time constraint considered here[6]. Note also that the limiting cases $|\nu|\to\infty$ correspond to the degenerate case

[6]Such an extension will become relevant when enrichening the SSA-criterion with additional timeliness or lead-lag requirements, not shown here.

$\mathbf{b} \propto \boldsymbol{\gamma}_\delta$ since then $\boldsymbol{\nu}/\nu \to -\mathbf{I}$ (the 'correct' sign can be accommodated by D). Equivalently, the difference-equation 21 morphs into an identity, up to arbitrary scaling. Another interesting limiting case occurs when the structure of the prediction problem is such that $\mathbf{b}$ approaches one of the eigenvectors $\mathbf{v}_i$ of $\mathbf{M}$, denoted by $\mathbf{b} \to \mathbf{v}_i$. Then $D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}\mathbf{b} \to (2\lambda_i - \nu)\mathbf{v}_i$. If $\boldsymbol{\gamma}_\delta$ is a fixed target with complete spectral support, then we conclude that $|D| \to 0$, since otherwise $D\boldsymbol{\gamma}_\delta \to (2\lambda_i - \nu)\mathbf{v}_i$ would imply $w_j = 0$ for $j \neq i$. If $|D| \to 0$ then, equivalently, $|\lambda_2| \to \infty$ which would imply $i = 1$ or $i = L$ and $|\rho_1| \to \rho_{max}(L)$[7]. On the other hand, if $\boldsymbol{\gamma}_\delta$ is not fixed and is allowed to approach $\mathbf{v}_i$ too, denoted by $\boldsymbol{\gamma}_\delta \to \mathbf{v}_i$, and if $|D| \not\to 0$ then $D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}\mathbf{b} \to (2\lambda_i - \nu)\mathbf{b}$ so that $\mathbf{b} \propto \boldsymbol{\gamma}_\delta$ (degenerate case) and $\rho_1 \to \lambda_i$. We conclude that if $\mathbf{b} \to \mathbf{v}_i$ then either $i \in \{1, L\}$ and $|\rho_1| \to \rho_{max}(L)$ (boundary cases of admissibility) or $\mathbf{b} \propto \boldsymbol{\gamma}_\delta \to \mathbf{v}_i$ and $\rho_1 \to \lambda_i$ for $i \in \{2, ..., L-1\}$ (asymptotically singular degenerate case with incomplete spectral support). For $|\nu| > 2\rho_{max}(L)$, the proof of assertion 5 implies that lag-one ACF $\rho\Big(y(\nu), y(\nu), 1\Big)$ and target correlation $\rho(y(\nu), z, \delta)$ of the SSA-solution $y(\nu)$ are linked by

$$\frac{\partial\rho\Big(y(\nu), z, \delta\Big)/\partial\nu}{\partial\rho\Big(y(\nu), y(\nu), 1\Big)/\partial\nu} \quad = \quad -\text{sign}(\nu)\frac{1}{\left(\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta\right)^{3/2}\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}$$

where the quotient is well defined due to strict monotonicity. This equality posits a fundamental tradeoff or dilemma of target correlation and lag-one ACF for the SSA-solution: for $\nu \in \{x|x > 2\rho_{max}(L)\}$ an increase of $\rho\Big(y(\nu), y(\nu), 1\Big)$ (holding-time) necessarily means a decrease of $\rho\Big(y(\nu), z, \delta\Big)$ (or an increase of the MSE); and inversely for $\nu \in \{x|x < -2\rho_{max}(L)\}$. Assertion 4 also implies that $y(\nu)$ is smoother or unsmoother than the MSE-predictor $\boldsymbol{\gamma}_\delta$ depending on the sign of ν being positive or negative for $\nu \in \{x||x| > 2\rho_{max}(L)\}$.

The case of incomplete spectral support is addressed formally in the following corollary.

Corollary 1 *Let all regularity assumptions of the previous theorem hold except completeness so that $NZ \subset \{1, ..., L\}$ or, stated otherwise, there exists i_0 such that $w_{i_0} = 0$ in 9. Then:*

1. *For $\nu \in \mathbb{R} - \{2\lambda_i | i = 1, ..., L\}$ the functional form of the SSA-estimate is*

$$\mathbf{b}(\nu) = D\sum_{i \in NZ}\frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \tag{28}$$

with corresponding lag-one acf

$$\rho(\nu) = \frac{\sum_{i \in NZ}\frac{\lambda_i w_i^2}{(2\lambda_i - \nu)^2}}{\sum_{i \in NZ}\frac{w_i^2}{(2\lambda_i - \nu)^2}} =: \frac{M_1}{M_2} \tag{29}$$

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

2. *Let $\nu = \nu_{i_0} := 2\lambda_{i_0}$ where $i_0 \notin NZ$ with adjoined rank-defficient $\boldsymbol{\nu}_{i_0} = 2\mathbf{M} - \nu_{i_0}\mathbf{I}$. 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, the functional form of $\mathbf{b}(\nu_{i_0})$ can be 'spectrally completed' as in*

$$\mathbf{b}_{i_0}(\tilde{N}_{i_0}) := \mathbf{b}(\nu_{i_0}) + D\tilde{N}_{i_0}\mathbf{v}_{i_0} \tag{30}$$

with lag-one acf

$$\rho_{i_0}(\tilde{N}_{i_0}) = \frac{M_{i_0 1} + \lambda_{i_0}\tilde{N}_{i_0}^2}{M_{i_0 2} + \tilde{N}_{i_0}^2} \tag{31}$$

[7]Indeed, $|\lambda_2| \to \infty$ if and only if the objective function of the SSA-criterion is overruled by the holding-time constraint which happens if and only if $|\rho_1| \to \rho_{max}(L)$ (limiting cases of admissibility). The 'only if' part follows from the fact that if $|\rho_1| \not\to \rho_{max}(L)$, then the intersection of unit-sphere and holding-time hyperbola is a proper subspace of $\mathbb{R}^{L-2}$ so that λ_2 is bounded.

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

$$\tilde{N}_{i_0} = \pm\sqrt{\frac{\rho_1 M_{i_0 2} - M_{i_0 1}}{\lambda_{i_0} - \rho_1}} \tag{32}$$

ensures compliance with the holding-time 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. *Any ρ_1 such that $|\rho_1| < \rho_{max}(L)$ is admissible in the holding-time constraint.*

Proof

The first assertion follows directly from the Lagrangian equation 19

$$D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta = \mathbf{b}(\nu)$$

where $\boldsymbol{\nu}$ has full rank if $\nu \in \mathbb{R} - \{2\lambda_i | i = 1, ..., L\}$, as assumed. Under the case posited in the second assertion $\boldsymbol{\nu}_{i_0}$ does not have full rank anymore and $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ as defined by 30 is a solution of the Lagrangian equation

$$D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}_{i_0}\mathbf{b}_{i_0}(\tilde{N}_{i_0})$$

for arbitrary $\tilde{N}_{i_0}$ since now $\mathbf{v}_{i_0}$ belongs to the kernel of $\boldsymbol{\nu}_{i_0}$. Moreover, orthogonality of $\mathbf{V}$ implies that

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

Solving for the holding-time constraint $\rho_{i_0}(\tilde{N}_{i_0}) = \rho_1$ then leads to

$$N_{i_0} := \tilde{N}_{i_0}^2 = \frac{\rho_1 M_{i_0 2} - M_{i_0 1}}{\lambda_{i_0} - \rho_1}$$

We infer that N_{i_0} is always positive if $0 < \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} < \rho_1 < \lambda_{i_0}$ or $0 > \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} > \rho_1 > \lambda_{i_0}$, so that $\tilde{N}_{i_0} = \pm\sqrt{N_{i_0}} \in \mathbb{R}$, 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 virtue of the intermediate-value theorem we then infer that any ρ_1 such that $|\rho_1| < \rho_{max}(L)$ is admissible for the holding-time constraint. Otherwise, if $w_1 = 0$ then $\mathbf{b}_1(\tilde{N}_1)$, where $i_0 = 1$ in 30, can 'fill the gap' and reach out the lower boundary $-\rho_{max}(L)$ as $\tilde{N}_1 \to \infty$. A similar reasoning would apply in the case $w_L = 0$ which achieves the proof of the corollary.

The case of incomplete spectral support is illustrated by a fleshed-out example in section 4.9 and we now proceed to the derivation of a numerical optimization routine for determining $\mathbf{b}$ in 6.

Corollary 2 *Let the assumptions of theorem 1 hold. Then the solution to the SSA-optimization problem 6 is*

$$\mathbf{b}(\nu_0) = sign_{\nu_0}\sum_{i=1}^{L}\frac{w_i}{2\lambda_i - \nu_0}\mathbf{v}_i \tag{33}$$

where ν_0 is a solution to the non-linear equation

$$\frac{\mathbf{b}(\nu_\mathbf{0})'\mathbf{M}\mathbf{b}(\nu_\mathbf{0})}{\mathbf{b}(\nu_\mathbf{0})'\mathbf{b}(\nu_\mathbf{0})} = \rho_1 \tag{34}$$

If the search for an optimal ν can be restricted to $|\nu| \geq 2\rho_{max}(L)$, then the solution to 34 is unique. Otherwise, the SSA-solution is determined by that particular solution ν_0 of 34 which maximizes the absolute value of the SSA objective function. Finally, the sign $sign_{\nu_0} = \pm 1$ is selected such that $\mathbf{b}(\nu_\mathbf{0})'\boldsymbol{\gamma}_\delta > 0$ leading to a positive objective function and criterion value.

A proof follows readily from assertions 2-4 of theorem 1. The distribution of the SSA-predictor is derived in the following corollary.

Corollary 3 *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 $\hat{\mathbf{b}}$ are*

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

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

The proof readily follows from 10. We here refer to standard textbooks for a derivation of mean, variance and (asymptotic) distribution of the MSE-estimate under various assumptions about x_t, see e.g. Brockwell and Davis (1993). The next corollary derives a dual interpretation of the SSA-predictor.

Corollary 4 *Let all regularity assumptions of Theorem (1) hold and let $y_t(\nu_0)$ denote the SSA-solution for some $\nu_0 > 2\rho_{max}(L)$. Set $\rho_{\nu_0,\delta} := \rho(y(\nu_0), z, \delta) > 0$ and consider the dual optimization problem*

$$\left.\begin{array}{c} \max_\mathbf{b} \rho(y, y, 1) \\ \rho(y, z, \delta) = \rho_{\nu_0,\delta} \end{array}\right\} \tag{35}$$

A solution to this problem has the same functional form $\mathbf{b} = \mathbf{b}(\nu)$ as in equation 10 and if the search for ν can be restricted to the set $\{\nu||\nu| > 2\rho_{max}(L)\}$ then $y_t(\nu_0)$ is also the solution of the dual problem. Similarly, if $\nu_0 < -2\rho_{max}(L)$, then $y_t(\nu_0)$ is also the solution of the dual problem

$$\begin{array}{c} \min_\mathbf{b} \rho(y, y, 1) \\ \rho(y, z, \delta) = \rho_{\nu_0,\delta} \end{array}$$

where maximization has been replaced by minimization.

Proof: The Lagrangian Equation 18 does not discern constraint and objective: after suitable re-scaling of Lagrange multipliers the problem specified by criterion 35 leads to the same functional form $\mathbf{b} = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta$ of its solution (re-scaling of Lagrange multipliers is always possible because the regularity assumptions of theorem 1 imply 'activated' constraints with non-vanishing and finitely-sized multipliers). The only difference to the original criterion 6 is that ν in 35 must be selected such that $\rho(y(\nu), z, \delta) = \rho_{\nu_0,\delta}$. If the search can be restricted to $\{\nu|\nu > 2\rho_{max}(L)\}$, then 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 $\{\nu||\nu| > 2\rho_{max}(L)\}$ in the corollary is readily obtained by considering that $\rho(y(\nu), y(\nu), 1)$ for $\nu \in \{x|x < -2\rho_{max}(L)\}$ is always smaller than $\rho(y(\nu), y(\nu), 1)$ if $\nu \in \{x|x > 2\rho_{max}(L)\}$, by assertion 4 of Theorem 1. The proof in the case $\nu_0 < -2\rho_{max}(L)$ is similar. In particular, minimization is necessary because $\rho(y(\nu), y(\nu), 1)$ for $\nu \in \{x|x < -2\rho_{max}(L)\}$ is always smaller (therefore minimization) than $\rho(y(\nu), y(\nu), 1)$ if $\nu \in \{x|x > 2\rho_{max}(L)\}$.

We shall see below that the restriction $\{\nu||\nu| > 2\rho_{max}(L)\}$ in the above corollary is not a limitation in a typical prediction context. Before, the follwing proposition analyzes the case $\nu \in [-2, 2]$.

Proposition 6 *Let all regularity assumptions of Theorem* (1) *hold, assume* $\nu \in]-2,2[$ *and let* $\mathbf{b}(\nu) = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta$. *Consider the solution* $\tilde{\mathbf{b}}$ *of the difference equation*

$$\tilde{b}_{k+1} \quad = \quad \nu\tilde{b}_k - \tilde{b}_{k-1} + D\gamma_{k+\delta} \ , \ 0 \le k \le L-1 \tag{36}$$

with arbitrary initializations. Then $\mathbf{b}(\nu)$ *with coefficients* $b_k = b_k(\nu)$ *is given by*

$$b_k = \tilde{b}_k + C\cos(\phi + \arccos(\nu/2)k) \tag{37}$$

where C, ϕ *can be chosen such that* $b_{-1} = b_L = 0$. *If* $\nu = 2$ *then* $\mathbf{b}(\nu)$ *is given by*

$$b_k = \tilde{b}_k + b + ak \tag{38}$$

and if $\nu = -2$, *then*

$$b_k = \tilde{b}_k + c + a(-1)^k k \tag{39}$$

where a, c *can be selected such that* $b_{-1} = b_L = 0$.

Proof: According to theorem 1 the coefficients $b_k = b_k(\nu)$ of $\mathbf{b}(\nu)$ must conform to the same difference equation as $\tilde{\mathbf{b}}$, subject to $b_{-1} = b_L = 0$. If $\tilde{b}_{-1} = \tilde{b}_L = 0$ then $\mathbf{b}(\nu) = \tilde{\mathbf{b}}$ and $C = 0$ in 37 ($a = b = 0$ in 38 or 39). Otherwise, the difference $\mathbf{db} := \mathbf{b}(\nu) - \tilde{\mathbf{b}}$ must be a solution of the homogeneous difference equation

$$db_{k+1} - \nu db_k + db_{k-1} = 0 \ , \ 0 \le k \le L-1$$

If $\nu \in]-2,2[$ then $db_k = C\cos(\phi + \arccos(\nu/2)k)$ where ϕ, C can be selected such that $b_{-1} = b_L = 0$. If $\nu = 2$ or $\nu = -2$ then the solutions of the homogeneous difference equations are $c + ak$ and $c + a(-1)^k k$, respectively, and a, c can be selected such that $b_{-1} = b_L = 0$.

In a prediction framework, typically, γ_k and b_k should decay towards zero for increasing lag k such that the remote past of a time series becomes increasingly irrelevant for determining its future: we refer to this property as the tacit prediction premise. If γ_k is such that $\sum_{k=-\infty}^{\infty} |k\gamma_k| < \infty$, then the MA-inversion of the AR(2)-filter with parameters $a_1 = -\nu, a_2 = 1$ determining $\tilde{b}_k$ in 36 is convergent (the AR(2) has either isolated single unit roots, if $|\nu| \neq 2$, or double roots, if $|\nu| = 2$ which are cancelled by γ_k). Since the MA-inversion is a solution of 36 there exists an initialization $\tilde{b}_0, \tilde{b}_1$ of $\tilde{\mathbf{b}}$ such that $\tilde{\mathbf{b}}$ corresponds to the inversion. The following proposition then suggests that the case $\nu \in [-2,2]$ contradicts the tacit prediction premise under fairly general assumptions.

Proposition 7 *Let all regularity assumptions of Theorem* (1) *hold, assume* $\nu \in [-2,2]$, $\sum_{k=-\infty}^{\infty} |k\gamma_k| < \infty$ *and assume, also, that* $\tilde{\mathbf{b}}(\nu, L)$ *in 36 is the MA-inversion of the AR(2)-equation. Then* $\tilde{b}_{-1}(\nu,\infty) := \lim_{L\to\infty} \tilde{b}_{-1}(\nu, L)$ *exists. If* $\tilde{b}_{-1}(\nu,\infty) \neq 0$, *then the coefficients* $b_k(\nu, L)$ *of* $\mathbf{b}(\nu, L)$ *do not converge towards zero for increasing lag* k *and for arbitrarily large* $L \to \infty$.

Proof: Under the above assumptions, the MA-inversion is convergent and therefore $\tilde{b}_{-1}(\nu,\infty) := \lim_{L\to\infty} \tilde{b}_{-1}(\nu, L)$ exists; also, $\lim_{L\to\infty} \tilde{b}_L(\nu, L) = 0$. If $\nu = 2$ then $c(\nu, L), a(\nu, L)$ in 38 or 39 must converge since they depend on $\tilde{b}_{-1}(\nu, L)$ and $\tilde{b}_L(\nu, L)$ only. Then

$$\lim_{L\to\infty} b_k = \lim_{L\to\infty} \tilde{b}_k + c(\nu, L) + a(\nu, L)k \to_{k\to\infty} c(\nu,\infty) + a(\nu,\infty)k = -\tilde{b}_{-1}(\nu,\infty) \neq 0$$

A similar proof applies to the case $\nu = -2$. If $|\nu| < 2$ is fixed, then $C(\nu, L), \phi(\nu, L)$ in 37 must converge (since they depend on $\tilde{b}_{-1}(\nu, L)$ and $\tilde{b}_L(\nu, L)$ only) and

$$\lim_{L\to\infty} b_k(\nu, L) = \lim_{L\to\infty} \tilde{b}_k(\nu, L) + C(\nu, L)\cos(\phi(\nu, L) + \arccos(\nu/2)k) \to_{k\to\infty} C(\nu,\infty)\cos(\phi(\nu,\infty) + \arccos(\nu/2)k)$$

where $C(\nu,\infty) \neq 0$ and $\arccos(\nu/2) \neq 0$. Finally, if $\nu = \nu(L) < 2$ is such that $\lim_{L\to\infty} \nu(L) = 2$, then

$$\lim_{L\to\infty} b_k(\nu, L) = \lim_{L\to\infty} \tilde{b}_k(\nu, L) + C(\nu, L)\cos(\phi(\nu, L) + \arccos(\nu(L)/2)k) \to_{k\to\infty} C(\nu,\infty)\cos(\phi(\nu,\infty))) = -\tilde{b}_{-1}(\nu,\infty) \neq 0$$

The condition $\sum_{k=-\infty}^{\infty}|k\gamma_k|<\infty$ ensures that the MA-inversion $\tilde{b}_k(\nu,L)$ exists and converges for all ν and that $\lim_{k\to\infty}\tilde{b}_k(\nu,L)=0$ so that $|\tilde{b}_k|$ decay towards zero. If $|\nu|\neq 2$, then the weaker absolute summability $\sum_{k=-\infty}^{\infty}|\gamma_k|<\infty$ is sufficient for ensuring existence and convergence of the MA-inversion. In a BCA-framework, typically, the target coefficients γ_k converge exponentially fast to zero so that the above condition is fulfilled; also, $\tilde{b}_{-1}(\nu,\infty)\neq 0$[8]. Interestingly, proposition 7 shows that $\mathbf{b}(\nu,L)$ is subject to a unit-root and its parameters do not decay to zero if $|\nu|<2$ due to the boundary constraint $b_{-1}=0$, at least if $\tilde{b}_{-1}(\nu\infty)\neq 0$. This outcome is to some extent paradoxical, because the boundary constraints $b_{-1}=b_L=0$ ensure boundary stability of the solution $\mathbf{b}$ of the unstable SSA AR(2) difference equation. We conclude that for $|\nu|<2$, boundary stability together with stability of the MA-inversion are not sufficient to ensure overall stability of $\mathbf{b}(\nu,L)$.
Under the above assumptions, the variance of the target z_t exists but if $|\nu|<2$ then the variance of $y_t(\nu)$ diverges for increasing L (the parameters do not decay to zero) so that the MSE diverges, too. Therefore, $y_t(\nu)$ cannot be a solution to the SSA-problem if $|\nu|\leq 2$ and if L is sufficiently large. We then deduce that the restriction $|\nu|>2\rho_{max}(L)$ in corollary 4 is not a limitation for large L since the more stringent $|\nu|>2$ applies. However, the unit root case $|\nu|<2$ cannot be excluded in all situations. If, for example, ρ_1 is close to $\rho_{max}(L)$ so that the holding-time ht_1 in the constraint is (too) large when compared to L, then the unit-root case might apply, due to strong smoothing requirements. In such a case, we argue that the prediction problem is ill-posed and increasing L or decreasing ht_1 will address the problem. To avoid such an ill-conditioned case, proposition 7 emphasized large $L\to\infty$.
The case $|\nu|>2$ differs in the sense that the roots of the AR(2) polynomial are not unit-roots anymore but real reciprocals λ and $1/\lambda$ with $\lambda\in\mathbb{R}\setminus\{0\}$ and $\nu=\lambda+1/\lambda$. In this case the solution of the homogeneous difference equation is $C_1(\lambda,L)\lambda^k+C_2(\lambda,L)\lambda^{L-k}$ and

$$b_k(\lambda,L)=\tilde{b}_k(\lambda,L)+C_1(\lambda,L)\lambda^k+C_2(\lambda,L)\lambda^{L-k}$$

where the parameters $C_1(\lambda,L),C_2(\lambda,L)$ are selected such that $b_{-1}(\lambda,L)=b_L(\lambda,L)=0$. Under the above assumptions $C_1(\lambda,L),C_2(\lambda,L)$ converge to $C_1(\lambda,\infty),C_2(\lambda,\infty)$ which depend on $\tilde{b}_1(\nu,\infty)$ and $\tilde{b}_\infty(\nu,\infty)=0$. If $\tilde{b}_k(\lambda,L)$ is the MA-inversion, then

$$0=\lim_{L\to\infty}b_L(\lambda,L)=\lim_{L\to\infty}\tilde{b}_L(\lambda,L)+C_1(\lambda,L)\lambda^L+C_2(\lambda,L)=C_2(\lambda,\infty)$$

and therefore $C_2(\lambda,\infty)=0$ so that the unstable root $1/\lambda$ asymptotically vanishes. Similarly, the effect of the homogeneous equation, namely $C_1(\lambda,L)\lambda^k$, vanishes exponentially fast and therefore $\lim_{L\to\infty}b_k(\lambda,L)$ can converge to zero with increasing lag k, in contrast to the unit-root case of proposition 7.

The above arguments were derived in a time-domain perspective, analyzing solutions of difference equations. We can derive analogous results in the frequency-domain. The SSA AR(2)-filter in equation 36 has amplitude function

$$A_{SSA}(\omega)=\frac{1}{\exp(-i\omega)-\nu+\exp(i\omega)}=\frac{1}{2\cos(\omega)-\nu}$$

Expression 23 for the SSA-solution $\mathbf{b}(\nu)$ is the exact discrete frequency-domain representation of the convolution of AR(2)-filter and $\boldsymbol{\gamma}_\delta$, recognizing that $\lambda_i=\cos(\omega_i)$ with Fourier frequencies ω_i. We can now distinguish the cases $\nu\leq -2$, $-2<\nu<2$ and $\nu\geq 2$. In the first case, $\nu\leq -2$, $A_{SSA}(\omega)$ is a highpass with a peak at frequency π; for $\nu\to\infty$ the filter becomes an allpass,

[8]The condition $\tilde{b}_{-1}(\nu,\infty)=0$ implies that the MA-inversion is compliant with the boundary constraints, without correction by the solution of the homogeneous equation, and this property does not relate to 'structures' such as imposed by classic prediction or signal extraction applications. In any case, corresponding $\boldsymbol{\gamma}_\delta$ generally look 'odd', in particular if $\gamma_k\geq 0$ is monotonically decreasing with k.

assuming suitable scaling by D, so that $\mathbf{b}(\nu) \to \boldsymbol{\gamma}_\delta$, confirming previous findings. The highpass leads to noise leakage requested when ht_1 in the holding-time constraint is smaller than ht_{MSE}. If $\nu = -2$ then the singularity of $A(\omega)$ at $\omega = \pi$ is of order two. For $\nu \geq 2$ the AR(2)-filter is a lowpass: emphasizing low frequencies leads to smoothing required when $ht_1 > ht_{MSE}$. For $-2 < \nu < 2$, the AR(2)-filter is a bandpass. The bandpass can be pertinent for example if $\boldsymbol{\gamma}_\delta$ is a strongly periodic sequence with slowly decaying (or non-decaying: unit-root) coefficients (narrow spectral peak) and if ht_1 in the holding-time constraint is close to the half-periodicity[9] of $\boldsymbol{\gamma}_\delta$, a rather atypical prediction problem in economic applications. Note, however, that 23 is subject to multiple singularities and that the lag-one ACF $\rho(y(\nu), y(\nu), 1)$ in 24 is no more monotonous such that multiple (up to L) different solutions to equation 34 are possible. Also, numerical computations can be very difficult due to densely packed singularities at $\nu = \lambda_i$, $i = 1, ..., L$ in $[-2, 2]$.

4 An Illustration of Technical Features

Our examples in this section address specific methodological features of the SSA-predictor: introductory forecast exercises are proposed in section 4.1; elements of signal extraction are examined in section 4.2; ; smoothing and 'un-smoothing' nowcasters are proposed in section 4.3; a filtering perspective is offered in section 4.4 with frequency-domain convolution, plancherel-identity and amplitude functions; a unit-root case is discussed in section 4.5; resilience against departures from Gaussianity is explored in section 4.6; section 4.7 presents a smoothness-timeliness dilemma and a discussion of SSA hyper-parameters; section 4.8 highlights multiplicity and uniqueness results; finally, the singular case of a target with incomplete spectral support is illustrated in section 4.9. SSA-solutions are based on corollary 2 and an open-source R-package is at disposal for replication, see (insert link to Github).

4.1 Forecasting

4.1.1 MA-Process

We consider a simple introductory forecast exercise for a MA(2)-process

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

with forecast horizon $\delta = 1$ (one-step ahead). We can set-up the problem in our general framework

$$z_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k} \ , \ x_t = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k} \ , \ \gamma_k = \begin{cases} 1 \ , & k = 0 \\ 0 \ , & \text{otherwise} \end{cases} \ , \ \xi_k = \begin{cases} 1 \ , & k = 0, 1, 2 \\ 0 \ , & \text{otherwise} \end{cases}$$

For comparison purposes, we compute three different SSA-predictors $y_{ti}, i = 1, 2, 3$ for z_t: the first two are of identical length $L = 20$ with dissimilar holding-times ht =3.74 and 10; the third predictor deviates from the second one by selecting $L = 50$; the holding-time of the first predictor matches the lag-one autocorrelation of z_t and is obtained by inserting $\rho(z, z, 1) = 2/3$ into 3. In addition, we also consider the MSE forecast $\hat{z}_{t,1}^{MSE} = \epsilon_t + \epsilon_{t-1}$, as obtained by classic time series analysis, as well as a trivial 'lag-by-one' forecast $\hat{z}_{t,1}^{lag\ 1} = z_t$, see fig. 1 (an arbitrary scaling scheme is applied to SSA filters). Predictors based on the 'true' MA(2)-model of z_t are virtually indistinguishable from predictors based on a fitted empirical model, see table 2 below.

[9] If ht_1 is larger than the half-periodicity, stronger smoothing may ask for a lowpass AR(2)-filter with $\nu > 2$, and similarly for smaller ht_1.

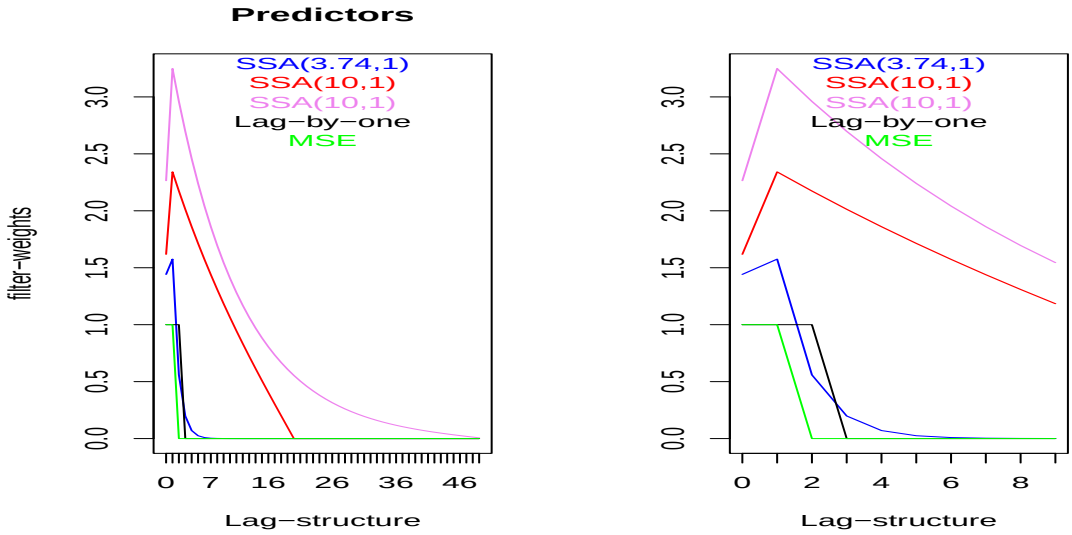

Figure 1: MSE-, SSA- and lag-by-one predictors with arbitrarily scaled SSA-designs. All lags (left panel) and first ten lags (right panel).

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

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

Table 1: Performances of MSE and lag-by-one benchmarks vs. SSA: All filters are applied to a sample of length 1000000 of Gaussian noise. Empirical holding-times are obtained by dividing the sample-length by the number of zero-crossings.

other forecasts in terms of correlation and sign accuracy but it loses in terms of smoothness or holding-time; SSA(3.74,1) outperforms the lag-by-one benchmark; both SSA(10,1) loose in terms of sign-accuracy but win in terms of smoothness and while the profiles of longer and shorter filters differ in figure 1, their respective performances are virtually indistinguishable in table 1, suggesting that the selection of L is not critical (assuming it is at least twice the holding-time). The table also illustrates the tradeoff between MSE- or sign-accuracy performances of optimal designs, in the top and bottom rows, and smoothing-performances in the middle row (an explicit formal link can be obtained but is omitted here). Finally, table 2 displays results when all predictors rely on an empirical model fitted to z_t on a data-sample of length 50: a comparison of both tables suggests that performances are virtually unaffected by the additional estimation step. To conclude, note that if $\delta > 2$, then $z_{t+\delta}$ is independent of $\epsilon_t, \epsilon_{t-1}, ...$ and thus of any predictor $y_t = \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$, contradicting thereby the first regularity assumption of theorem 1 (identifiability). Formally, the

	SSA(3.88,1)	SSA(10,1)	SSA(10,1)	Lag-by-one	MSE
Correlation with target	0.774	0.386	0.388	0.680	0.815
Empirical holding-times	3.885	10.000	10.000	3.885	2.987
Empirical sign accuracy	0.782	0.626	0.627	0.738	0.803

Table 2: Same case as above but all predictors rely on an empirical model of the MA(2)-process: the model is fitted on a sample of length 50.

objective function or criterion value $\rho(y, z, \delta) = 0$ vanishes for all $\mathbf{b}$ so that the SSA-predictor is not identified anymore: all predictors are equally valid or invalid. The best MSE-forecast $\hat{z}_{t+\delta} = 0$ is well-defined, though.

4.1.2 AR(1)-Process

We here consider a more challenging forecast exercise for the AR(1)-process

$$z_t = -0.9 z_{t-1} + \epsilon_t$$

with forecast horizon $\delta = 1$ (one-step ahead). The regular sign-alternating pattern of its acf (or of its realizations) is a salient feature of this process whose holding-time $ht_1 = 1.17$ approaches the lower limit $ht_1 = 1$. The specific challenge of this exercise consists in deriving a 'smooth' SSA-predictor with $ht_1 = 5$, tracking z_{t+1} as closely as possible i.e. y_t must reconcile two strongly conflicting requirements. We first note that

$$z_t \quad = \quad \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k} \ , \ x_t = \sum_{k=0}^{\infty} \xi_k \epsilon_{t-k}$$

with either

$$\gamma_k = \left\{ \begin{array}{cc} 0 & , \ k < 0 \\ (-0.9)^k & , \ k \geq 0 \end{array} \right. \ , \ \xi_k = \left\{ \begin{array}{cc} 1 & , \ k = 0 \\ 0 & , \ \text{otherwise} \end{array} \right.$$

or

$$\gamma_k = \left\{ \begin{array}{cc} 1 & k = 0 \\ 0 & \text{otherwise} \end{array} \right. \ , \ \xi_k = (-0.9)^k, k \geq 0$$

The latter formulation relies on the extension to autocorrelated x_t, see section 2.3. Figure 1 displays and compares MSE- and SSA-predictors (an arbitrary scaling scheme is applied to the latter). Specifically, the convolution $(\mathbf{b} \cdot \boldsymbol{\xi})$ of the generalized SSA-criterion 7 can be obtained directly from theorem 1: the corresponding predictor-weights are applied to the MA-inversion of the AR(1) i.e. to ϵ_{t-k}, see the bottom panels of the figure. In this case, the MSE-predictor is $\sum_{k=0}^{\infty}(-0.9)^{k+1}\epsilon_{t-k}$ with weights $b_k^{MSE} = (-0.9)^{k+1}, k \geq 0$ (green). The SSA predictor-weights as assigned to the proper AR(1)-process z_{t-k} can be obtained by deconvolution 8, as displayed in the top-panels of the figure: in this case the MSE-predictor is $-0.9 z_t$ with weight -0.9 at lag zero and zero otherwise (green line). The bottom-left panel suggests that the alternating MSE-pattern (green) is carried over to SSA (blue) whose coefficients are additionally 'lifted away' from the zero-line: in combination with the alternating pattern the resulting hybrid profile ensures conformity with the imposed holding-time constraint while maximizing correlation with the target $z_{t+\delta} = z_{t+1}$.

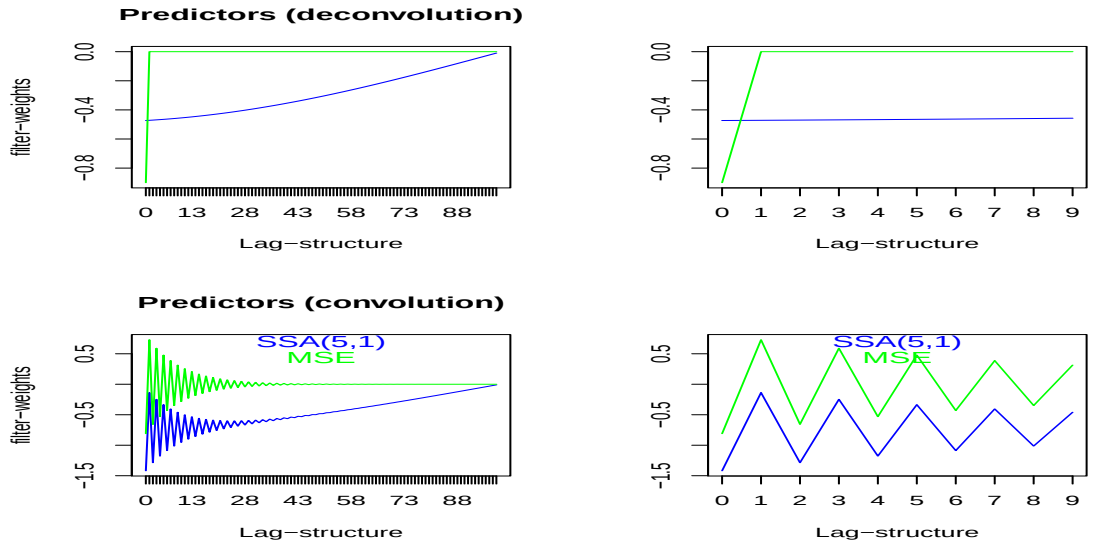

Figure 2: SSA- and MSE- (one step ahead) forecasts. All lags (left panel) and first ten lags (right panel). Predictor weights as assigned to z_{t-k} (top: deconvolution) and to ϵ_{t-k} (bottom: convolution).

Table 2 compares predictor performances: SSA maximizes correlation or sign-accuracy with the target z_{t+1} subject to $ht_1 = 5$. Finally, fig.3 compares the target z_{t+1} and its SSA-predictor y_t: maximization of the cross-correlation of both series is obtained by synchronization of the corresponding alternating high-frequency patterns; at the same time, the marked low-frequency swings of SSA (blue) about the zero line ensure conformity with the holding-time constraint.

	SSA(5,1)	MSE
Correlation with target	0.293	0.900
Empirical holding-times	5.000	1.168
Empirical sign accuracy	0.595	0.856

Table 3: Performances of SSA- and MSE-predictors.

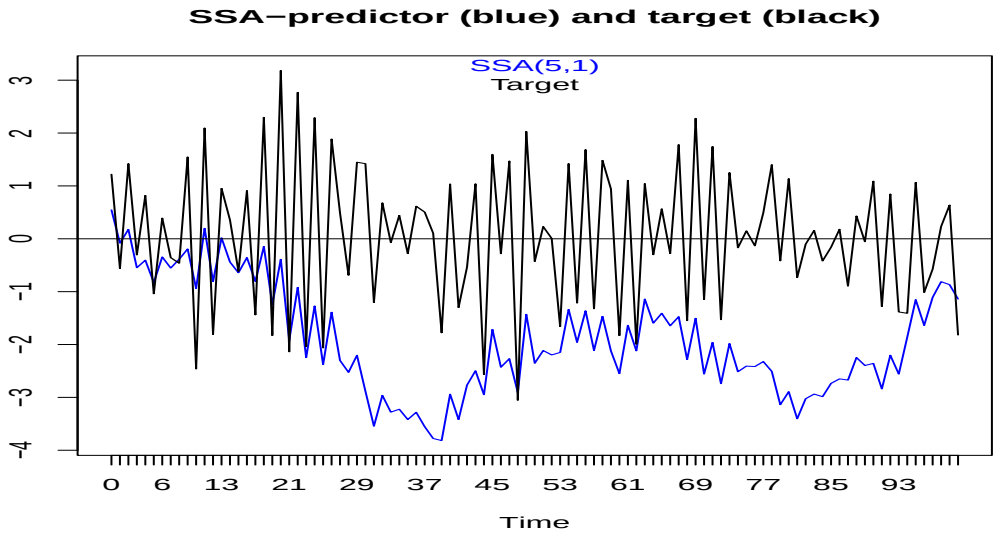

Figure 3: SSA-predictor (blue) and target $z_{t+\delta}$ (black).

4.2 Elements of Signal Extraction

We consider one-step ahead forecasting of the slightly more complex target

$$z_t = x_t + x_{t-1} + x_{t-2} \tag{40}$$
$$x_t = -0.3x_{t-1} + \epsilon_t + 0.7\epsilon_{t-1} + 0.8\epsilon_{t-2} \tag{41}$$

where x_t is a stationary ARMA(1,2)-process with MA-inversion or Wold-decomposition $x_t = \sum_{k\geq 0}\xi_k\epsilon_{t-k}$ where $\xi_k = \left(\mathbf{A}_{ar}^k\mathbf{b}_{ma}\right)_1$, $\mathbf{A}_{ar} = \begin{pmatrix} -0.3 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$, $\mathbf{b}_{ma} = (1, 0.7, 0.8)'$ and where $(\cdot)_1$ denotes the first element of a vector[10]. This example can be related to signal extraction, where a target filter $\boldsymbol{\gamma}$ (the equally-weighted MA(3) in 40) is applied to autocorrelated data x_t (the ARMA-process in 41) in order to extract interesting components: Wildi (2023) relies on a bi-infinite symmetric Hodrick-Prescott design, see Hodrick and Prescott (1997), as the target or 'signal extraction' filter $\boldsymbol{\gamma}$ for the analysis of business-cycles; the equally-weighted MA(3) filter $\gamma_0 = \gamma_1 = \gamma_2 = 1$ in 40 is a correspondingly much simpler lowpass filter. SSA-predictors in this framework can be set-up in terms of $y_t = \sum_{k=0}^{L-1}(b \cdot \xi)_k\epsilon_{t-k}$ or $y_t = \sum_{k=0}^{L-1} b_k x_{t-k}$: the latter corresponds to the proper signal extraction filter or predictor. The convolution $(\mathbf{b} \cdot \boldsymbol{\xi})$, proposed in section 2.3, can be obtained directly from theorem 1, see the bottom panels of fig.4 whereby we impose ht_1 =3.74 and length L =20 (blue), ht_1 =10, L =20 (red) and ht_1 =10, L =50 (violet).

[10]Invertibility of the ARMA is not required for deriving the SSA-predictor: theorem 1 assumes that $\boldsymbol{\gamma}_\delta$ has complete spectral support which applies here.

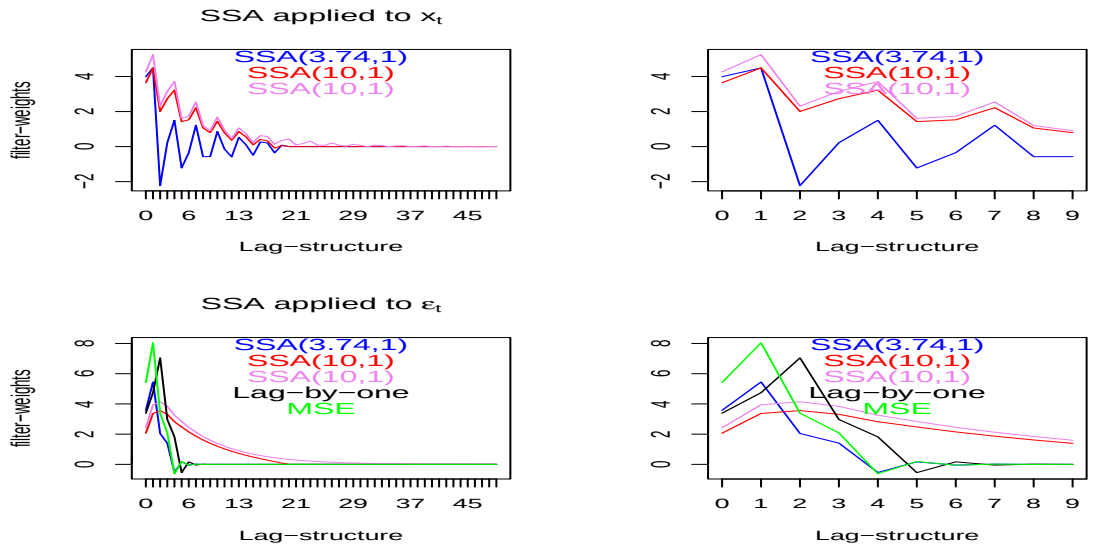

Figure 4: SSA arbitrarily scaled. All lags (left panel) and first ten lags (right panel). Predictors as applied to x_t (upper panels) and ϵ_t (bottom panels).

The signal extraction filter $\mathbf{b}$, displayed in the top panels, can be obtained by inversion or deconvolution of $(b \cdot \xi)_j$, see 8. To conclude, note that the selection of L is generally uncritical as long as $L > 2ht_1$.

4.3 Smoothing and Un-Smoothing

In this example we aim at fitting a time series z_t conditional on different holding-times: in contrast to the previous two sections we here emphasize *nowcasting* i.e. $\delta = 0$. For this purpose, we consider the target

$$\begin{aligned} z_t &= x_t + x_{t-1} + x_{t-2} \\ x_t &= 0.8x_{t-1} + \epsilon_t + 0.5\epsilon_{t-1} + 0.4\epsilon_{t-2} \end{aligned}$$

with holding-time ht =10.4. We then apply two SSA-designs with holding-times 3.7 and 30: the first nowcast 'un-smooths', the second smooths z_t, whilst minimizing MSE (up to arbitrary scaling). Fig.5 displays filter coefficients as applied to x_t, left panel, or to ϵ_t, right panel (arbitrarily scaled). Note the typical shape of the filters in the right panel, indicating presence of the left boundary constraint $(b \cdot \xi)_{-1} = 0$ (the filters in the left panel are subject to different boundary-constraints, due to deconvolution).

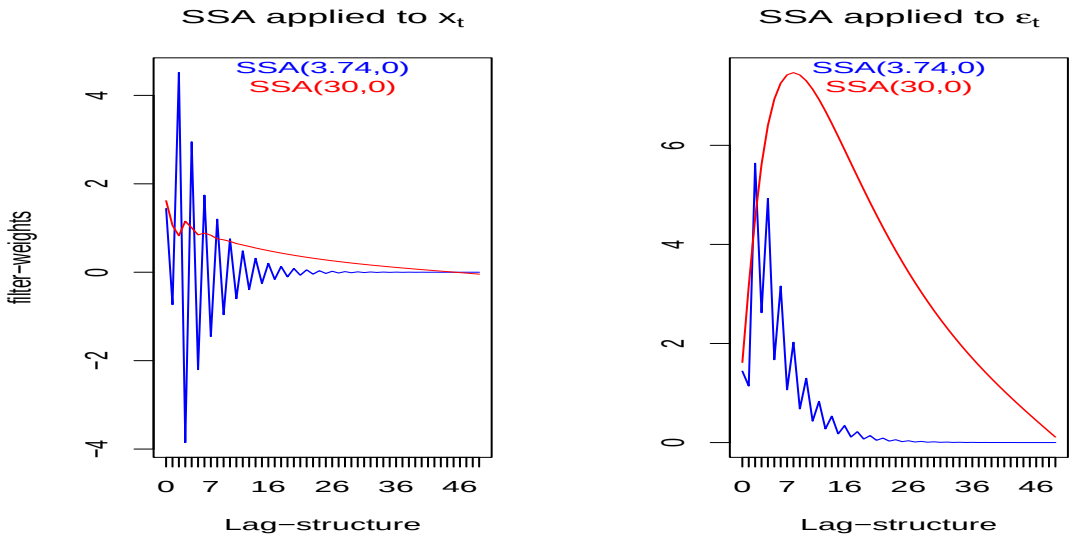

Figure 5: SSA arbitrarily scaled: smoothing nowcaster (red) and un-smoothing nowcaster (blue), as applied to x_t (left) and ϵ_t (right).

SSA-nowcasters and target are compared in fig.6: all series are arbitrarily scaled to unit-variance. Zero-crossings of the smoothing nowcast (red) are fewest, followed by the target (black) and the un-smoothing nowcast (blue): frequencies or occurrences of crossings are inversely proportional to holding-times, as desired. The maximized theoretical criterion values $\rho(y_{SSA_i}, z, 0)$, $i = 1, 2$, are 0.928 (blue un-smoother) and 0.648 (red smoother): these match empirical estimates 0.932 and 0.653 based on a sample of length 20000 of (Gaussian) x_t from which an excerpt is shown in fig.6.

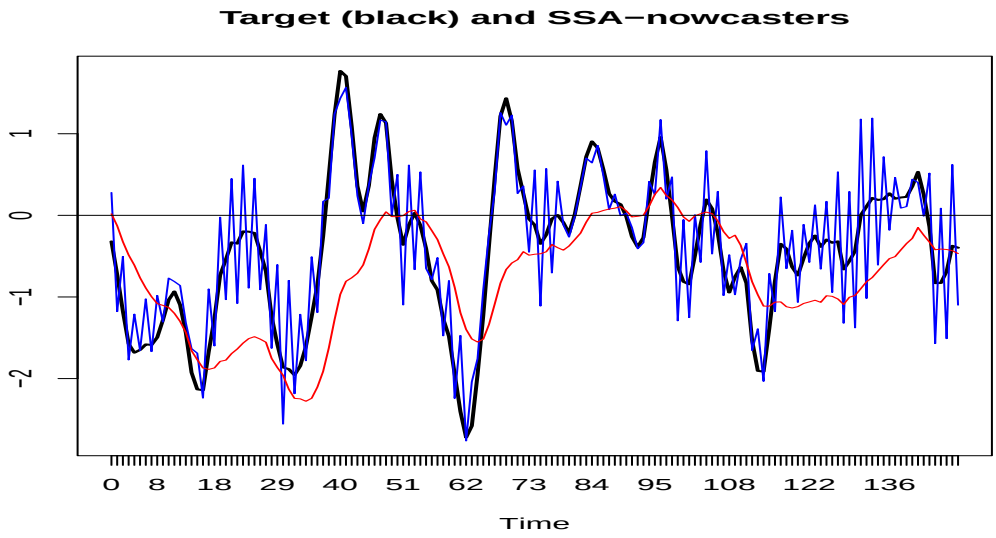

Figure 6: Target (black) and SSA-nowcasts (blue and red): all series scaled to unit-variance.

4.4 Filtering

According to the extension in section 2.3 and to theorem 1, the SSA-solution $(\mathbf{b}\cdot\boldsymbol{\xi})$ can be obtained from 21 as

$$(b\cdot\xi)_{k+1} - \nu(b\cdot\xi)_k + (b\cdot\xi)_{k-1} = D\gamma_{k+\delta} \tag{42}$$

In principle, this equation suggests that $(\mathbf{b}\cdot\boldsymbol{\xi})$ can be interpreted as the time-domain convolution of a 'generic SSA' AR(2)-filter, with AR-coefficients $a_1 = -\nu$, $a_2 = 1$, and the (scaled) MSE-filter $D\boldsymbol{\gamma}_\delta$. While an exact formal treatment of the time-domain solution, as entailed by 42, must be deferred, due to lengthier technical features related among others to the instability of the AR(2)[11] we can nevertheless refer to the equivalent frequency-domain expression 10

$$(\mathbf{b}\cdot\boldsymbol{\xi}) = D\sum_{k=1}^{L}\frac{w_k}{2\lambda_k-\nu}\mathbf{v}_k = D\sum_{k=1}^{L}\frac{w_k}{2(-\cos(\omega_k))-\nu}\mathbf{v}_k = D\sum_{k=1}^{L}\frac{w_k}{(-\exp(i\omega_k))-\nu+(-\exp(-i\omega_k))}\mathbf{v}_k$$

where we inserted $-\cos(\omega_k) = \lambda_k$ computed at the Fourier-frequencies $\omega_k = k\pi/(L+1)$ for the eigenvalues λ_k of $\mathbf{M}$, see section 3 and Anderson (1975). The rightmost term formalizes the convolution in the frequency-domain of generic SSA AR(2)- and MSE-filters, whereby the latter is represented by its spectral weights w_k; the amplitude function of the generic SSA AR(2) is $\frac{1}{|-\exp(i\omega_k)-\nu-\exp(-i\omega_k)|} = \frac{1}{|2\lambda_k-\nu|}$; moreover, 11 reflects the Plancherel-identity. Remarkably, these standard 'stationary' frequency-domain results apply despite instability of the generic SSA AR(2)-filter. Fig.7 displays amplitude functions of both SSA-nowcasts of the previous section 4.3: SSA-amplitudes in left and right panels correspond to SSA-nowcasts in left and right panels of fig.5. The additional violet line in fig.7 is the amplitude of the ARMA-filter ξ_k, $k \geq 0$, and the amplitude functions in the left panel $A_{SSA,x_t}(\omega)$ can be obtained by division of SSA- by ARMA-amplitudes in the right panel $A_{SSA,x_t}(\omega) = A_{SSA,\epsilon_t}(\omega)/A_{Arma}(\omega)$: this frequency-domain deconvolution corresponds to the time-domain deconvolution 8. The 'un-smoothing' nowcast applied to x_t (blue line, left panel of fig.7) is a highpass filter, emphasizing high-frequency-components of x_t: as a result the number of zero-crossings increases (blue vs. black line in fig.6). The smoothing nowcast (red line, left panel fig.7) is a lowpass filter: as a result zero-crossings are fewer (red vs. black line in fig.6).

[11] The additional boundary constraints $b_{-1} = b_L = 0$ are main determinants of the solution, ensuring 'stability' in the presence of the potentially unstable AR-equation.

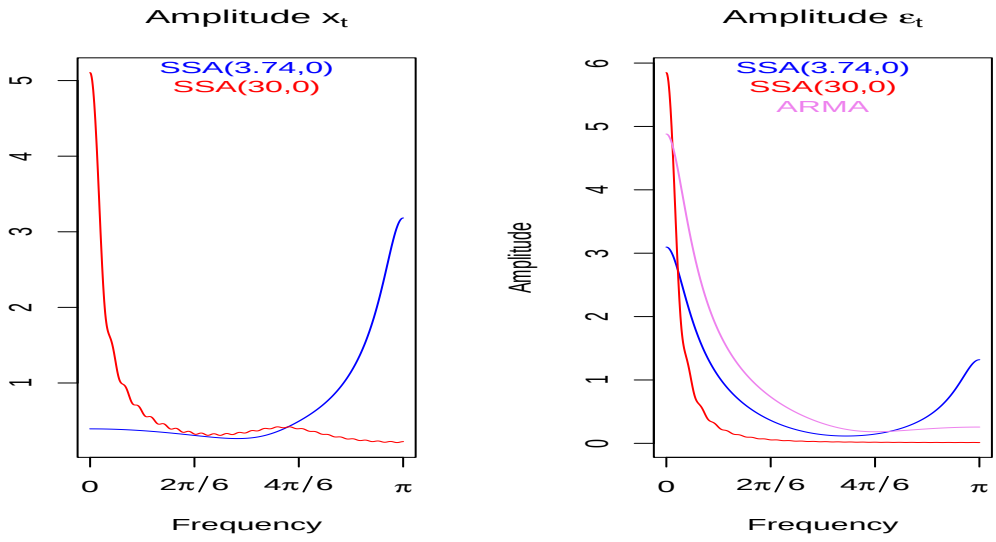

Figure 7: Amplitude functions of SSA-nowcaster as applied to x_t (left panel) and to ϵ_t(right panel) with arbitrary scaling. The amplitude of the ARMA-filter x_t is displayed in the right panel (violet)

4.5 Unit-Root Case

Typically, the coefficients of a SSA-predictor $\mathbf{b}$ decay towards zero at a rate determined by $\boldsymbol{\gamma}_\delta$ and the holding-time constraint ρ_1 or ht_1. Figure 8 displays two SSA-designs for the nowcast problem in section 4.3: both nowcasters are subject to the same holding-time $ht_1 = 20$ but they differ in terms of filter-lengths: $L = 20$ (blue) vs. $L = 100$ (red). The shorter filter is subject to a unit-root, since $\nu_0 = 1.992 < 2$, where ν_0 is the (numerical) solution to the holding-time equation 34. As a result the forecast weights do not follow an exponential-law at higher lags: the pattern is close to the eigenvector $\mathbf{v}_{20}$ of the largest eigenvalue $\lambda_{20} = \rho_{max}(20) = 0.9888$ of $\mathbf{M}$ (upper half of a sinusoid). Indeed, the imposed holding-time $ht = 20$ approaches the upper limit of admissibility $ht_{max}(20) = 21$ for a MA-filter of length $L = 20$, see proposition 3, so that 'smoothing' starts to supplant 'forecasting' and that $\mathbf{b} \to \text{sign}(w_L)\mathbf{v}_L = \mathbf{v}_{20}$, see first claim of theorem 1. Increasing the filter length L from 20 to 100 reinstates stability as illustrated by the exponential decay of the 'long' SSA-nowcaster at higher lags (red line, right panel). In applications, the unit-root case typically occurs when L and ht_1 are mismatched. In such a case, an increase of the filter-length improves MSE-performances by unleashing degrees of freedom that were previously frozen: in our example the criterion value $\rho(y, z, \delta)$ rises ceteris paribus from 0.72, for the short nowcaster (left panel), to 0.77, for the long nowcaster (right panel). Once again, the choice of L is generally uncritical, provided $L > 2ht_1$.

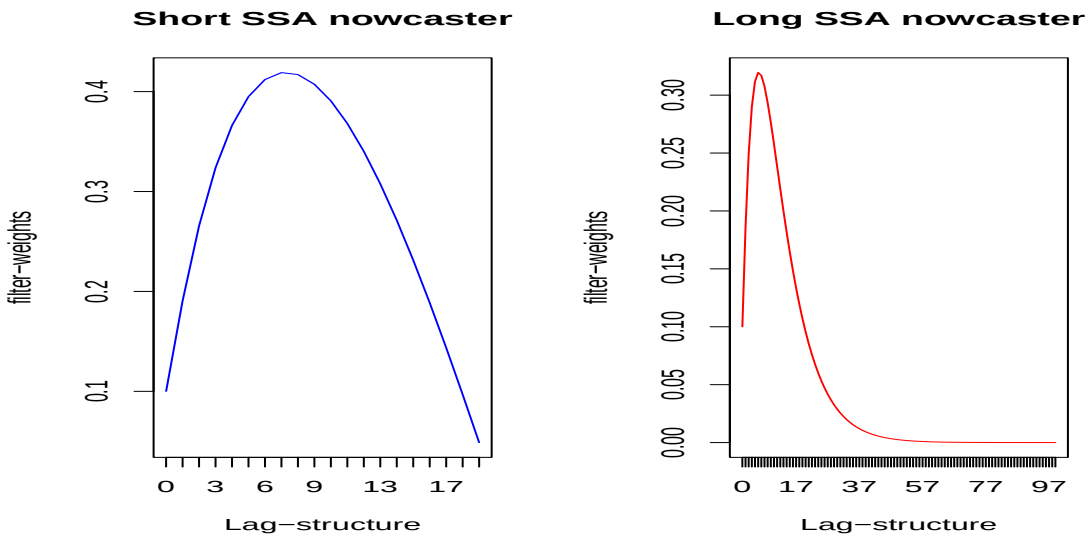

Figure 8: Arbitrarily scaled SSA nowcasters for identical holding-time ht=20 but different filter lengths L=20 (blue) and L=100 (red).

4.6 Resilience

Wildi (2023) applies the SSA-approach to (the log-returns of) a broadly diversified equity index, the Standard and Poors 500, as well as to industrial production indices of a selection of countries with long and consistent sample histories. Empirical and theoretical holding-times (the latter wrongly assume Gaussianity) match virtually perfectly for the financial time series, despite volatility clustering, non-vanishing mean (drift) and extreme observations during financial and pandemic crises. Discrepancies observed in the case of the macro-indicators were attributable to autocorrelation and could be alleviated by the extension to autocorrelated processes illustrated in the above sections. We here complement these empirical findings, which document resilience of the approach against departures from Gaussianity, by an application of SSA to white noise $x_t = \epsilon_t$ where ϵ_t is t-distributed with degrees of freedom ranging from $df = 2$ (heavy tails) to $df = 10$ (nearly Gaussian). We then compare empirical holding-times, i.e. length of filter-outputs divided by number of zero-crossings, to theoretical holding-times, wrongly assuming Gaussianity, based on long samples of size 100000 of ϵ_t, see table 4. Our findings suggest that an increased incidence of

	SSA(3.74,1)	SSA(10,1):L=20	SSA(10,1):L=50
t-dist.: df=2	3.77	11.14	11.26
t-dist.: df=4	3.78	10.45	10.54
t-dist.: df=6	3.78	10.32	10.42
t-dist.: df=8	3.77	10.14	10.22
t-dist.: df=10	3.76	10.18	10.18

Table 4: Empirical holding-times of SSA designs as applied to t-distributed white noise

extreme observations (first and second rows of the table) leads to a positive bias of the empirical holding-times for filters with a larger ht or lag-one acf. This phenomenon can be explained by the impulse response of the filters which is triggered by extreme observations and which does not change sign because all filter coefficients are positive: a longer tail of the filter then implies fewer crossings and a positive bias of the empirical holding-time. But the magnitude of the bias seems to be well controlled, overall, even in the presence of series with heavy-tails and the bias could be

reduced further by application of outlier techniques (not shown here).

4.7 A Smoothness-Timeliness Dilemma

Often, stronger noise-rejection or smoothing by an optimal (nowcast or forecast) filter is associated with increased lag or 'right-shift' of its output: the following example illustrates that the mentioned tradeoff, a so-called smoothness-timeliness dilemma, does not hold in general. For illustration, we rely on a simple empirical framework where the target $z_t := \frac{1}{100}\sum_{k=0}^{99}\epsilon_{t-k}$ is the output of an equally-weighted MA-filter of length L =100 applied to simulated Gaussian noise ϵ_t. The target must be forecasted at the horizon 20 by a classic MSE as well as by a SSA(30,20)-filter, whose holding-time $ht = 30$ exceeds that of the MSE design $ht = 19.8$ by a safe margin. Out of curiosity, we also supply a second SSA(30,40)-filter optimized for forecast horizon 40: the two hyperparameters ht, δ of the two SSA-designs suggest that for an identical smoothing capability or holding-time, the second filter should have improved timeliness properties in terms of a lead or left-shift. The three (arbitrarily scaled) forecast filters are displayed in fig.9[12] and filter outputs, arbitrarily scaled to unit-variance, are compared in fig.10.

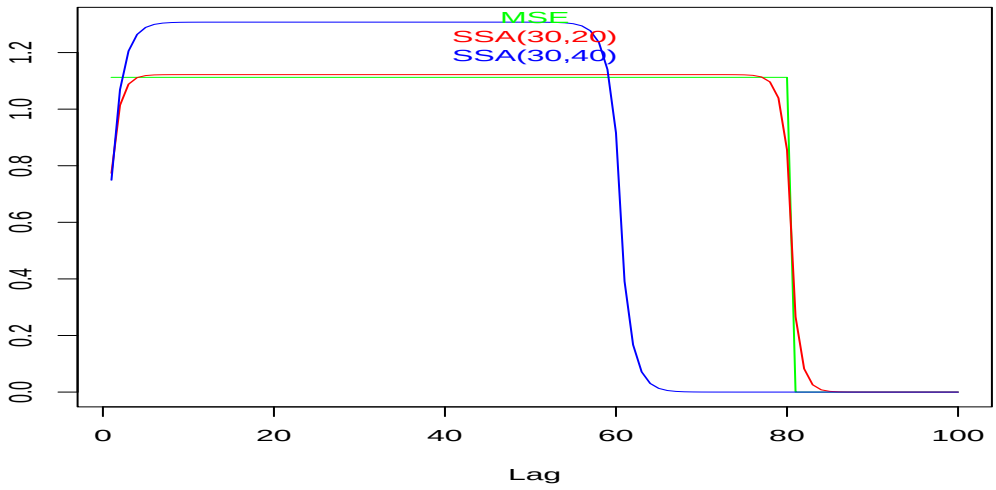

Figure 9: Forecast filters: MSE (green), SSA(30,20) (red) and SSA(30,40) (blue) with arbitrary scaling

[12]The early rise at the left edge reveals the presence of the left-side boundary constraint $b_{-1} = 0$, recall theorem 1.

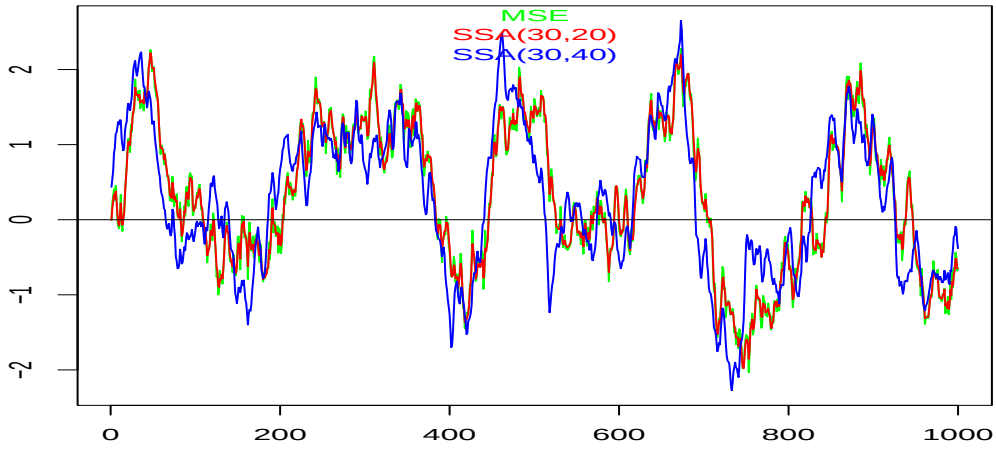

Figure 10: Outputs of forecast filters: MSE (green), SSA(30,20) (red) and SSA(30,40) (blue)

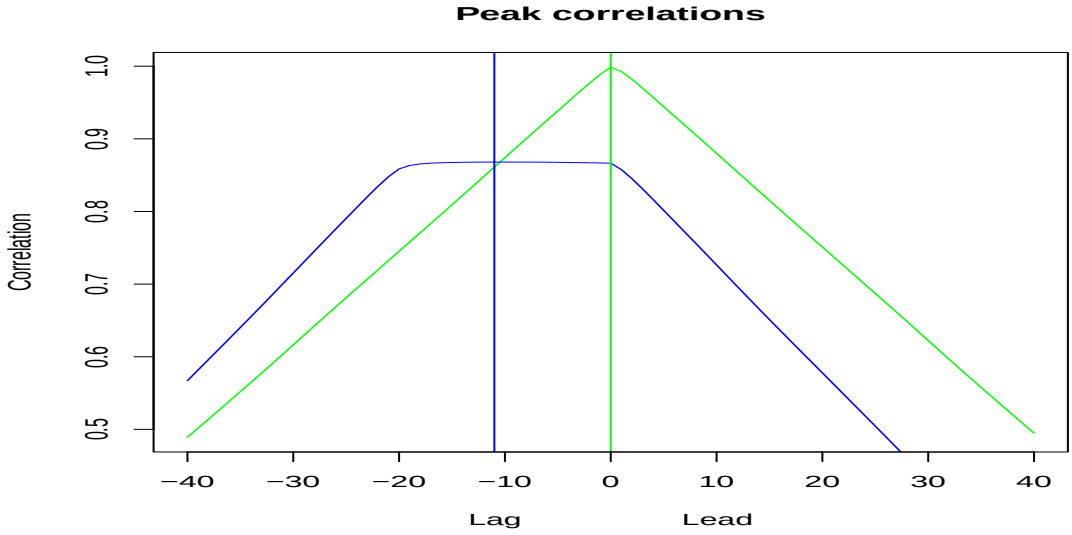

Figure 11: Correlation of shifted SSA(30,20) vs. MSE (green) and SSA(30,40) (blue). Positive numbers correspond to a relative lead of SSA(30,20) over the contenders. Peak correlations are indicated by vertical lines.

As expected, the output of SSA(30,40) (blue line in fig.10) appears left-shifted. Fig.11 displays cross-correlations at various leads and lags of the reference SSA(30,20): the relative shift can be inferred from the peak-correlation i.e. the lead or lag at which the maximum is achieved. The figure suggests that SSA(30,20) and MSE are on par (green line) and that SSA(30,20) lags or, equivalently, that SSA(30,40) leads by 11 time-units (blue line). Finally, the empirical holding-times in table 5, computed on a sample of length 100000, conform to expected values, as based on 3. We conclude that for identical smoothing capabilities, SSA(30,40) has improved timeliness

	MSE	SSA(30,20)	SSA(30,40)
1	19.9	30.9	30.9

Table 5: Empirical holding-times of MSE and SSA designs

characteristics in terms of a systematic lead; moreover, SSA(30,40) outperforms MSE in terms of timeliness and smoothness; also, timeliness and smoothness can be addressed explicitly by specifying hyper-parameters (ρ_1, δ). In this abstract context, the pair (ρ_1, δ) spans a two-dimensional space of predictors SSA(ρ_1, δ), for a particular target $z_{t+\delta_0}$, with distinct smoothness and timeliness characteristics entailed by the hyper-parameters: we argue that ρ_1, δ can be selected in view of matching particular research priorities, see e.g. Wildi (2023). Classic MSE-performances can be replicated by selecting $\delta = \delta_0$ and $\rho_1 = \rho_{MSE}$, the lag-one acf of the mean-square predictor.

4.8 Monotonicity vs. Non-Monotonicity

We here illustrate uniqueness or multiplicity of the solution of the non-linear holding-time equation 34, depending on $|\nu| > 2\rho_{max}(L)$ or $|\nu| \leq 2\rho_{max}(L)$, see assertion 4 of theorem 1. Fig. 12 displays the lag-one autocorrelation $\rho(\nu)$ in 11 for a SSA-nowcast ($\delta = 0$) as a function of ν for two different AR(1)-targets $\boldsymbol{\gamma}_0(a_1) = (1, a_1, ..., a_1^9)'$ of length $L = 10$ with $a_1 = 0.99$ (bottom panels) and $a_1 = 0.6$ (top panels). The panels on the left correspond to $|\nu| < 2\rho_{max}(10)$ and illustrate non-monotonicity of $\rho(\nu)$; the panels on the right correspond to $\nu > 2\rho_{max}(10)$ and illustrate strict monotonicity[13].

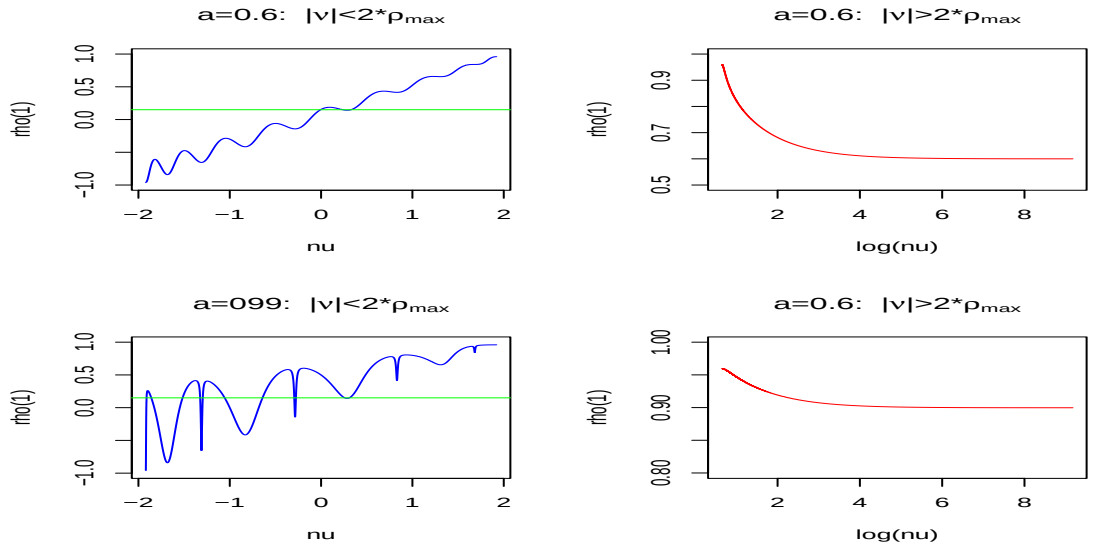

Figure 12: Lag-one autocorrelation as a function of ν when the target is a classic AR(1) with $a_1 = 0.6$ (top) and $a_1 = 0.99$ (bottom): the left/right-split of the panels corresponds to $|\nu| \leq 2\rho_{max}(L)$ (left) and $\nu > 2\rho_{max}(L)$ (right)

Non-monotonicity generally leads to multiple solutions $\nu_1, ..., \nu_n$ to the holding-time equation 34 for given ρ_1, whereby the multiplicity can depend on ρ_1, L as well as on the target $\gamma_{k+\delta}$: as can be seen in fig.12, the green horizontal line corresponding to $\rho_1 = 0.15$ intersects the acf (blue) four times in the upper left panel, meaning evidence of four different solutions $\nu_1, ..., \nu_4$ for given

[13]The abscissa of the right hand panels are based on transformed $\log(\nu)$ for a better visualisation of the monotonic shape.

$\rho_1 = 0.15$; in the bottom-left panel, the green line corresponding to $\rho_1 = 0.15$ intersects the acf (blue) $L+1 = 11$ times, meaning eleven different solutions $\nu_1, ..., \nu_{11}$ for this particular target. On the other hand, strict monotonicity of $\rho(\nu)$ as a function of ν in the right panels means that ν is determined uniquely by ρ_1. To conclude, in case of multiple solutions $\nu_1, ..., \nu_n$ to the holding-time equation 34, the SSA-solution $\mathbf{b}(\nu_{i_0})$ given by 33 is determined by that $\nu_{i_0} \in \{\nu_1, ..., \nu_n\}$ which maximizes the objective function of the SSA-criterion.

4.9 Incomplete Spectral Support

In order to illustrate the case of incomplete spectral support addressed by corollary 1 we here consider a simple nowcast example (forecast horizon $\delta = 0$) based on a band-limited target $\boldsymbol{\gamma}_0$ of length $L = 10$

$$\boldsymbol{\gamma}_0 = \sum_{i=1}^{10} w_i \mathbf{v}_i$$

where $\mathbf{v}_i$ are the eigenvectors of the $10 * 10$-dimensional autocovariance generating matrix $\mathbf{M}$ and where the last three weights in the spectral decomposition vanish, $w_8 = w_9 = w_{10} = 0$ ($m = 7$ in 9), and the first seven weights are constant $w_i = 0.378$, $i = 1, ..., 7$

$$\boldsymbol{\gamma}_0 = \sum_{i=1}^{7} 0.378 \mathbf{v}_i$$

The left panel in fig. 13 displays the lag-one acf 29 of $\mathbf{b}(\nu)$ given by 28 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 31 of the extension $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ in 30, when $\nu = \nu_{i_0} = 2\lambda_{i_0}$ for $i_0 = 8, 9, 10$, where the three additional (vertical black) spectral lines, corresponding to $\mathbf{v}_8, \mathbf{v}_9, \mathbf{v}_{10}$, 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 31, 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 acfs, marked by colored vertical lines in each panel, indicate potential solutions of the SSA-problem for the thusly specified holding-time 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).

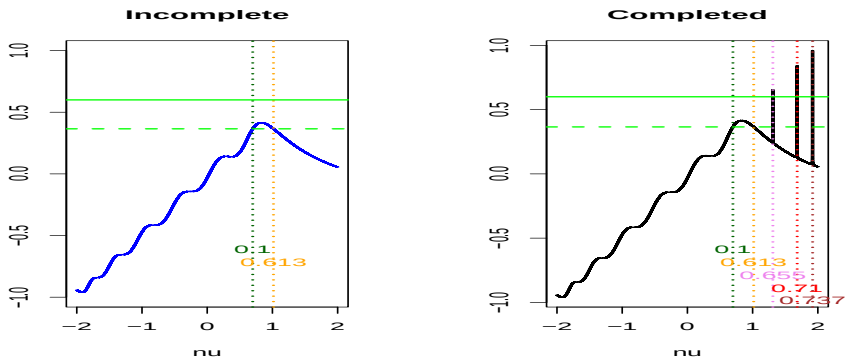

Figure 13: Lag-one autocorrelation as a function of ν. Original (incomplete) solutions (left panel) vs. completed solutions (right-panel). Intersections of the acf with the two green lines are potential solutions of the SSA-problem for the corresponding holding-times: criterion values are reported for 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 $i_0 = 8, 9, 10$ can accommodate for a wider range of holding-time constraints, such that $|\rho_1| < \rho_{max}(L) = \lambda_{10} = 0.959$; in contrast, $\mathbf{b}(\nu)$ in the left panel is limited to $-0.959 = \lambda_1 < \rho_1 < \lambda_7 = 0.415$ 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 holding-time 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

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

is attained at the right-most spectral line, for $i_0 = 10$, and where $\tilde{N}_{10} = 0.077$ has been obtained from 32, with the correct signs of D and $\tilde{N}_{10}$ in place.

5 Conclusion

We propose a novel SSA-criterion which emphasizes sign accuracy and zero-crossings of the predictor subject to a holding-time constraint. Under the Gaussian assumption, the classic MSE-criterion is equivalent to unconstrained SSA-optimization: in the absence of a holding-time constraint and down to an arbitrary scaling nuisance. We argue that the proposed concept is resilient against various departures from the Gaussian assumption. Moreover, the approach is interpretable and appealing due to its actual simplicity and because the criterion merges relevant facets of the prediction problem. While a formal treatment of timeliness, as an additional constitutional element of the prediction problem, would go beyond the scope of the proposed SSA-framework, our examples illustrate that alternative research priorities can be addressed consistently and effectively by a pair of hyper-parameters and the smoothness or holding-time constraint has a natural and interpretable meaning. Despite its structural simplicity, the predictor is feature-rich, as illustrated by reproducible examples of specific technical traits.

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] Brockwell P.J. and Davis R.A. (1993) Time Series: Theories and Methods (second edition). *Springer Verlag.*

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

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

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

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

[8] 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.

[9] McElroy, T. and Wildi , M. (2020) The multivariate linear prediction problem: model-based and direct filtering solutions. *Econometrics and Statistics* **14**, 112-130.

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

[11] Wildi, M. (2023) Business-Cycle Analysis and Zero-Crossings of Time Series: a Generalized Forecast Approach. Submitted for publication to the Journal of Business Cycle Analysis.