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

Marc Wildi

December 6, 2024

Abstract

Zero-crossings, or sign changes, in the stationary growth rate of a non-stationary time series indicate transitions between periods of expansion and contraction, which are essential for decision-making and control processes. In this framework, we present a multivariate extension of the recently developed Smooth Sign Accuracy (SSA) forecasting approach, termed M-SSA. This extension simultaneously addresses sign accuracy, mean squared error, and the frequency of zero-crossings in predictions. The proposed M-SSA criterion captures a fundamental dilemma by recognizing the trade-off between accuracy and smoothness (AS) of a predictor. As a result, the M-SSA framework accommodates a variety of design priorities related to AS forecasting performance, encompassing the conventional mean squared error paradigm. We explore the applications of this approach in three primary areas: forecasting, real-time signal extraction (nowcasting), and smoothing. In forecasting, M-SSA enables precise control over the zero-crossings of the predictor. The signal extraction example highlights the customization of AS forecasting performance within a traditional business-cycle analysis tool. Meanwhile, the smoothing aspect offers a simplified version of M-SSA, serving as either an alternative or a complement to established smoothing algorithms. Collectively, our applications illustrate the versatility and applicability of the M-SSA approach across a range of contexts.

Keywords: Forecasting, signal extraction, smoothing, zero-crossings, sign accuracy, mean-squared error, business-cycles.

1 Introduction

The analysis of zero-crossings in time series data was first introduced by Rice (1944), who established a correlation between the autocorrelation function (ACF) of a zero-mean stationary Gaussian process and the expected number of zero crossings within a specified time interval. In the context of non-stationary time series, sign changes in the growth rate signify transitions between phases of expansion and contraction. These alternating periods of growth and contraction can be linked to the business cycle, provided that the fluctuations exhibit sufficient magnitude and duration. To be classified as representative of a business cycle, consecutive zero-crossings of the growth rate must be separated by intervals potentially spanning several years. This requirement for duration necessitates a smooth trajectory of the corresponding economic indicator. The relationship between the smoothness of a time series and the mean duration between consecutive sign changes—termed the holding time (HT) by Wildi (2024)—is formalized in the work of Rice (1944).

Forecasting presents a complex estimation challenge, as it necessitates the consideration of various, often conflicting, factors. In this context, we emphasize two critical dimensions: accuracy and smoothness (AS). Accuracy pertains to the estimation of the future level of a time series, while smoothness is defined through the HT, which serves to regulate the frequency of undesirable 'noisy' sign changes. Ideally, these two dimensions could be jointly optimized, resulting in a predictor that closely approximates the true, albeit unobserved, future observation, thereby minimizing the incidence of spurious crossings, often referred to as 'false alarms.' However, the relationship between accuracy and smoothness presents a predictive dilemma; when appropriately formalized, an enhancement in one dimension inevitably leads to a compromise in the other. In this study, we propose a novel multivariate framework that enables users to manage and balance both accuracy and smoothness effectively. Furthermore, we extend existing methodologies to incorporate the concept of a HT, allowing for a customization of traditional 'benchmark' predictors in terms of AS performances.

The traditional mean-square error (MSE) forecast paradigm focuses primarily on accuracy, often at the expense of smoothness. In certain applications, this emphasis on accuracy can result in excessive noise leakage, leading to an increased incidence of false alarms, as demonstrated in Wildi (2025). To address this issue, McElroy and Wildi (2019) propose a forecasting approach based on a forecast trilemma; however, their solution does not take zero-crossings into account. Wildi (2024) introduces the SSA optimization criterion, which explicitly regulates the HT of a predictor through a smoothing constraint derived from Rice's foundational work. Additionally, Wildi (2025) proposes a formal treatment of the associated optimization problem with a strong emphasis on explainability. In this study, we extend this approach to multivariate prediction problems, emphasizing a stationary framework to elucidate the key concepts. An extension of the (univariate) SSA to non-stationary integrated processes has been discussed in Wildi (2025), who derives non-stationary maximal monotone and minimal curvature predictors, and these findings can be further extended to the M-SSA framework.

Our applications are designed to emphasize some of the distinctive characteristics of the M-SSA framework in forecasting, real-time signal extraction and smoothing. All examples can be replicated using an open-source M-SSA package, which includes an R package along with comprehensive instructions, practical use cases, and theoretical results, available at https://github.com/wiaidp/R-package-SSA-Predictor.git. A notable feature of our approach is the ability to customize benchmark predictors in terms of accuracy and smoothness (AS) performance. The M-SSA package demonstrates applications of this customization to various filtering techniques, including the Hodrick-Prescott (HP) filter (Hodrick and Prescott, 1997), Hamilton's regression filter (Hamilton, 2018), the Baxter-King (BK) bandpass filter (Baxter and King, 1999) and a 'refined' Beveridge-Nelson design (Kamber, Morley and Wong (2024)).

Section (2) briefly reviews the (univariate) SSA approach and introduces to the M-SSA crite-

rion; Section (3) proposes solutions to the optimization problem; Section (4) illustrates applications in forecasting, signal extraction and smoothing; finally, Section (5) summarizes our main findings.

2 M-SSA Criterion

We provide a brief review of the univariate case and the various adaptations of the Smooth Sign Accuracy (SSA) optimization criterion introduced by Wildi (2024). Following this review, we derive the multivariate generalization M-SSA.

2.1 Univariate SSA

We consider a zero-mean stationary time series denoted as x_t (processes with a non-vanishing mean are assumed to be centered) and a target variable $z_{t+\delta}$, where $\delta \in \mathbb{Z}$, which depends on future values of x_{t-k} for $k < 0$. We then derive an optimal predictor y_t for $z_{t+\delta}$ based on past values x_{t-k} for $k \geq 0$ (a causal filter). This predictor is characterized by the property that the HT of y_t—defined as the mean duration between consecutive zero-crossings—can be specified by the user. For clarity and simplicity, we initially assume $x_t = \epsilon_t$ to be an independent and identically distributed white noise (WN) sequence. An extension of this approach to accommodate autocorrelated stationary processes will be discussed subsequently.

Let $z_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k}$, where $x_j = \epsilon_j, j \in \mathbb{Z}$, is WN. Fot th sake of clarity we may assume that ϵ_t is standardized. The sequence $\boldsymbol{\gamma} = (\gamma_k)$, where $k \in \mathbb{Z}$, is a real, square summable sequence, ensuring that z_t constitutes a stationary process with zero mean and variance given by $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We seek to construct a predictor defined as $y_t = \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$ for the target $z_{t+\delta}$, where $\delta \in \mathbb{Z}$ and b_k represents the coefficients of a one-sided causal filter of length L. This predictive framework is commonly categorized as fore-casting, now-casting, or back-casting, depending on whether $\delta > 0$, $\delta = 0$, or $\delta < 0$, respectively. For illustration, consider the specific case where $\gamma_0 = 1$, $\gamma_1 = 0.5$, and $\gamma_k = 0$ for $k \notin 0, 1$. In this scenario, $z_t = \epsilon_t + 0.5\epsilon_{t-1}$ represents a moving-average process of order one. The task of one-step ahead forecasting for z_t can be achieved by setting $\delta = 1$. The classic mean squared error (MSE) estimate for z_{t+1} yields $y_t = 0.5\epsilon_t$, resulting in the coefficients $b_0 = 0.5$ and $b_k = 0$ for $1 \leq k \leq L - 1$. Furthermore, our target specification can also facilitate signal extraction problems, wherein the weights γ_k are coefficients of a two-sided (potentially bi-infinite) filter.

In general, the MSE predictor effectively captures the level of future observations (or targets) optimally; however, this approach may produce forecasts that are 'noisy'. This phenomenon is evident in the previous moving average (MA(1)) example, where $y_t = 0.5\epsilon_t$ exhibits white noise characteristics. To mitigate this issue, Wildi (2024, 2025) propose the following optimization problem:

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

Here, $\mathbf{b} = (b_0, ..., b_{L-1})'$ and $\boldsymbol{\gamma}_\delta = (\gamma_\delta, ..., \gamma_{\delta+L-1})'$ are column vectors of dimension L. The variable l represents a scaling parameter. The autocovariance generating matrix $\mathbf{M}$ is defined as:

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

which has dimension $L \times L$. The matrix $\mathbf{M}$ is structured such that $\mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{k=1}^{L-1} b_{k-1}b_k$, representing the lag-one autocovariance of y_t under the specified assumptions. This formulation

aims to optimize the predictor while controlling for the noise present in the forecasts. Criterion (1) is referred to as the SSA criterion, and its solution is denoted as $\text{SSA}(\rho_1, \delta)$, with the dependence on the scaling constant l omitted for brevity (see below for details). The constraints $\mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1$ and $\mathbf{b}'\mathbf{b} = l$ are termed the HT constraint and the length constraint, respectively. Under the assumption of WN, the classic MSE predictor is defined as $y_{t,MSE} := \boldsymbol{\gamma}_\delta' \boldsymbol{\epsilon}_t$, where $\boldsymbol{\epsilon}_t := (\epsilon_t, ..., \epsilon_{t-(L-1)})'$. The weights $\boldsymbol{\gamma}_\delta$ can be derived as a solution to the SSA criterion by setting $l := \boldsymbol{\gamma}_\delta' \boldsymbol{\gamma}_\delta$ and $\rho_1 := \boldsymbol{\gamma}_\delta' \mathbf{M} \boldsymbol{\gamma}_\delta / \boldsymbol{\gamma}_\delta' \boldsymbol{\gamma}_\delta =: \rho_{MSE}$. However, the objective is to ensure that the SSA predictor $y_t := \mathbf{b}' \boldsymbol{\epsilon}_t$ exhibits reduced noise characteristics compared to $y_{t,MSE}$ and the hyperparameter ρ_1 within the HT constraint can be utilized to this effect.

2.2   Alternative SSA Formulations

To streamline the terminology, we will consolidate the concepts of filter outputs and filter weights, thereby referring to both $y_{t,MSE}$ and $\boldsymbol{\gamma}_\delta$ as the MSE predictor. Similarly, we will merge y_t and $\mathbf{b}$ under the same designation, clarifying our intent when necessary. Under the assumption of white noise, $\boldsymbol{\gamma}\delta$ represents the appropriate target $\boldsymbol{\gamma}$ in Criterion (1), given that γ_k is irrelevant for $k < \delta$ or $k > \delta + L - 1$. We will also assume $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$. In this framework, the solution $\mathbf{b}_0$ to the SSA criterion can be interpreted as a constrained predictor for $z_{t+\delta}$. Alternatively, $\mathbf{b}_0$ can be viewed as a 'smoother' for $y_{t,MSE}$, as discussed by Wildi (2024). Thus, the SSA criterion simultaneously addresses prediction and smoothing. Moreover, under the imposed length constraint, the objective function $\mathbf{b}' \boldsymbol{\gamma}_\delta$ is proportional to:

$$\rho(y, z, \delta) := \mathbf{b}' \boldsymbol{\gamma}_\delta / \sqrt{l \boldsymbol{\gamma}' \boldsymbol{\gamma}}$$

which represents the target correlation of y_t with $z_{t+\delta}$, or to

$$\mathbf{b}' \boldsymbol{\gamma}_\delta / \sqrt{l \boldsymbol{\gamma}_\delta' \boldsymbol{\gamma}_\delta}$$

the correlation of y_t with $y_{t,MSE}$. Maximizing either of these functions consequently maximizes the other, allowing Criterion (1) to be reformulated as:

$$\left. \begin{array}{cc} \max_{\mathbf{b}} \rho(y, z, \delta) & \max_{\mathbf{b}} \rho(y, y_{MSE}, \delta) \\ \rho(y, y, 1) = \rho_1 \quad \text{or} & \rho(y, y, 1) = \rho_1 \\ \mathbf{b}'\mathbf{b} = l & \mathbf{b}'\mathbf{b} = l \end{array} \right\}. \tag{2}$$

Here, $\mathbf{b}'\mathbf{M}\mathbf{b}/l = \mathbf{b}'\mathbf{M}\mathbf{b}/\mathbf{b}'\mathbf{b} =: \rho(y, y, 1)$ denotes the lag-one autocorrelation function (ACF) of y_t. An increase in ρ_1 results in a stronger lag-one ACF for the predictor, thus yielding a 'smoother' trajectory for y_t characterized by less frequent zero-crossings.

Let $ht(y|\mathbf{b}, i)$ be defined as $E[t_i - t_{i-1}]$, where t_i, $i \geq 1$ represents the *consecutive* zero-crossings of the process y_t. Specifically, these crossings satisfy $t_{i-1} < t_i$ with the condition $t_1 \geq L$ and $\text{sign}(y_{t_{i-1}} y_{t_i}) < 0$ for all i. Additionally, it holds that $\text{sign}(y_{t-1} y_t) > 0$ if $t_{i-1} < t < t_i$. Under the assumptions of stationarity, $ht(y|\mathbf{b}, i) = ht(y|\mathbf{b})$ is invariant with respect to i. Furthermore, assuming Gaussianity, we have:

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

where this relationship remains relatively robust to deviations from Gaussian assumptions (Wildi, 2024). Given that the trigonometric function arccos is monotonic in the interval $]-1, 1[$, we can reformulate the SSA criterion as follows:

$$\left. \begin{array}{c} \max_{\mathbf{b}} \rho(y, z, \delta) \\ ht(y|\mathbf{b}) = ht_1 \\ \mathbf{b}'\mathbf{b} = l \end{array} \right\}. \tag{4}$$

4

This indicates a connection between the HT and the lag-one ACF. Notably, since correlations, signs, and zero-crossings are invariant to the scaling of y_t, the parameter l in the length constraint can be treated as a nuisance parameter, which primarily serves to ensure uniqueness. We typically assume $l = 1$ unless otherwise stated. If necessary, 'static' adjustments for level and scale can be applied post hoc, once a solution $\mathbf{b} = \mathbf{b}(l)$ has been determined for an arbitrary l. This adjustment can be performed, for instance, by regressing the predictor on the target variable. However, our primary focus lies in the 'dynamic' aspects of the prediction problem, particularly concerning the target correlation and sign accuracy (SA). Specifically, let $SA(y_t) = P(\text{sign}(z_{t+\delta}) = \text{sign}(y_t))$ denote the probability of the target and predictor sharing the same sign. Under the Gaussian assumption, it follows that:

$$SA(y_t) = 0.5 + \frac{\arcsin(\rho(y, z, \delta))}{\pi}$$

thereby establishing a strictly monotonic (bijective) relationship between SA and target correlation, which is also robust against deviations from Gaussianity (Wildi, 2024). Consequently, the SSA criterion can be reformulated as:

$$\max_{\mathbf{b}} SA(y_t)$$
$$ht(y|\mathbf{b}) = ht_1$$
$$\mathbf{b}'\mathbf{b} = 1,$$

In conclusion, the SSA framework effectively reconciles MSE, SA and HT, i.e. smoothing, in a flexible and interpretable manner.

2.3 Multivariate M-SSA

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{nt})'$ represent a multivariate stationary process of dimension n, characterized by a Wold decomposition of the form

$$\mathbf{x}_t = \sum_{k=0}^{\infty} \mathbf{\Xi}_k \boldsymbol{\epsilon}_{t-k},$$

where $\mathbf{\Xi}_0 = \mathbf{I}_{n \times n}$ denotes the identity matrix, and $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{nt})'$ is a sequence of multivariate WN with a variance-covariance matrix denoted as $\mathbf{\Sigma}$. The entries of $\mathbf{\Sigma}$ are represented by σ_{ij}, and we assume that its eigenvalues $\tilde{\sigma}_j$ for $j = 1, \ldots, n$ are strictly positive, indicating full rank. If $\mathbf{\Sigma}$ is originally rank deficient, we assume that redundant noise components have been eliminated, resulting in a lower-dimensional system that retains full rank. The eigenvectors of $\mathbf{\Sigma}$ are denoted by $\mathbf{v}_{\sigma k}$ for $k = 1, \ldots, n$. Next, we define a multivariate stationary process $\mathbf{z}_t$ as follows:

$$\mathbf{z}_t = \sum_{|k| < \infty} \mathbf{\Gamma}_k \mathbf{x}_{t-k},$$

where $\mathbf{\Gamma}_k$ is a sequence of square-summable filter matrices, with entries (γ_{ijk}) for $i, j = 1, \ldots, n$. The condition for square summability is expressed mathematically as:

$$\sum_{|k| < \infty} \text{tr}(\mathbf{\Gamma}_k' \mathbf{\Gamma}_k) < \infty,$$

where $\text{tr}(\cdot)$ denotes the trace operator (see Cerreia-Vioglio, 2023). We focus on the estimation of $\mathbf{z}_{t+\delta}$, utilizing the predictor defined by

$$\mathbf{y}_t = \sum_{k=0}^{L-1} \mathbf{B}_k \mathbf{x}_{t-k}.$$

To facilitate the theoretical analysis, we simplify our model by assuming $\mathbf{x}_t = \boldsymbol{\epsilon}_t$, with further extensions to autocorrelated processes addressed in the subsequent section. Let b_{ijk} denote the entries of the matrix $\mathbf{B}_k$ for $k = 0, \ldots, L-1$. We define several vector notations as follows:

$$\boldsymbol{\epsilon}_{it} = (\epsilon_{it}, \epsilon_{it-1}, ..., \epsilon_{it-(L-1)})' \quad , \quad \boldsymbol{\epsilon}_{\cdot t} = (\boldsymbol{\epsilon}'_{1t}, ..., \boldsymbol{\epsilon}'_{nt})'$$

$$\boldsymbol{\gamma}_{ij\delta} = (\gamma_{ij\delta}, \gamma_{ij\delta+1}, ..., \gamma_{ij\delta+L-1})' \quad , \quad \boldsymbol{\gamma}_{i\cdot\delta} = (\boldsymbol{\gamma}'_{i1\delta}, \boldsymbol{\gamma}'_{i2\delta}, ..., \boldsymbol{\gamma}'_{in\delta})'$$

$$\mathbf{b}_{ij} = (b_{ij0}, b_{ij1}, ..., b_{ijL-1})' \quad , \quad \mathbf{b}_i = (\mathbf{b}'_{i1}, \mathbf{b}'_{i2}, ..., \mathbf{b}'_{in})'$$

$$\boldsymbol{\gamma}_\delta = \begin{pmatrix} \boldsymbol{\gamma}'_{1\cdot\delta} \\ \boldsymbol{\gamma}'_{2\cdot\delta} \\ \cdot \\ \boldsymbol{\gamma}'_{n\cdot\delta} \end{pmatrix} \quad , \quad \mathbf{b} = \begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \cdot \\ \mathbf{b}'_n \end{pmatrix}.$$

We define $y_{ijt} := \mathbf{b}'_{ij}\boldsymbol{\epsilon}_{jt}$, which allows us to express $y_{it} = \mathbf{b}'_i \boldsymbol{\epsilon}_{\cdot t} = \sum_{j=1}^n y_{ijt}$ and $\mathbf{y}_t = \mathbf{b}'\boldsymbol{\epsilon}_{\cdot t}$. Through orthogonal projection, the MSE predictor is derived as $\hat{\mathbf{y}}_{t,MSE} = \boldsymbol{\gamma}'_\delta \boldsymbol{\epsilon}_{\cdot t}$. Consider the Kronecker products defined as $\tilde{\mathbf{I}} := \boldsymbol{\Sigma} \otimes \mathbf{I}_{L\times L}$ and $\tilde{\mathbf{M}} := \boldsymbol{\Sigma} \otimes \mathbf{M}$, of the $n \times n$ dimensional matrix $\boldsymbol{\Sigma}$, the $L \times L$ dimensional identity $\mathbf{I}_{L\times L}$ and the autocovariance generating matrix $\mathbf{M}$. We can express the following expectations:

$$
\begin{aligned}
E[z_{it+\delta}y_{it}] &= E[\hat{z}_{it\delta}y_{it}] = \boldsymbol{\gamma}'_{i\cdot\delta}\tilde{\mathbf{I}}\mathbf{b}_i \\
E[\hat{z}^2_{it\delta}] &= \boldsymbol{\gamma}'_{i\cdot\delta}\tilde{\mathbf{I}}\boldsymbol{\gamma}_{i\cdot\delta} \\
E[y^2_{it}] &= \mathbf{b}'_i\tilde{\mathbf{I}}\mathbf{b}_i \\
E[y_{it-1}y_{it}] &= \mathbf{b}'_i\tilde{\mathbf{M}}\mathbf{b}_i.
\end{aligned}
$$

The variance of z_{it} is given by

$$E[z^2_{it}] = \sum_{|k|<\infty} \boldsymbol{\gamma}'_{ik}\boldsymbol{\Sigma}\boldsymbol{\gamma}_{ik},$$

where $\boldsymbol{\gamma}_{ik} := (\gamma_{i1k}, ..., \gamma_{ink})'$. Consequntly, we propose the following multivariate M-SSA criterion for $i = 1, ..., n$:

$$
\begin{aligned}
\max_{\mathbf{b}_i} \boldsymbol{\gamma}'_{i\cdot\delta}\tilde{\mathbf{I}}\mathbf{b}_i & \quad (5) \\
\mathbf{b}'_i\tilde{\mathbf{M}}\mathbf{b}_i &= \rho_i \\
\mathbf{b}'_i\tilde{\mathbf{I}}\mathbf{b}_i &= 1.
\end{aligned}
$$

Under the imposed length constraint, the objective function is proportional to the correlation between the predictor and the i-th target, or equivalently, the i-th MSE predictor. Additionally, the HT constraint regulates the lag-one ACF or, equivalently, the HT of the predictor. The relationships established between sign accuracy and target correlation, as well as between lag-one ACF and HT, as discussed in the previous section, remain valid. Thus, the M-SSA framework can be reformulated in accordance with any alternative and equivalent expressions of the optimization criterion.

2.4 Dependence

Let us define the stationary processes as follows:

$$
\begin{aligned}
\mathbf{x}_t &= \sum_{m=0}^{\infty} \boldsymbol{\Xi}_m \boldsymbol{\epsilon}_{t-m} \\
\mathbf{z}_t &= \sum_{|k|<\infty} \boldsymbol{\Gamma}_k \mathbf{x}_{t-k}.
\end{aligned}
$$

Consequently, the target and predictor can be expressed formally as:

$$\mathbf{z}_t = \sum_{|k|<\infty} (\mathbf{\Gamma} \cdot \mathbf{\Xi})_k \boldsymbol{\epsilon}_{t-k}$$

$$\mathbf{y}_t = \sum_{j \geq 0} (\mathbf{B} \cdot \mathbf{\Xi})_j \boldsymbol{\epsilon}_{t-j}.$$

Here, $(\mathbf{\Gamma} \cdot \mathbf{\Xi})_k = \sum_{m \leq k} \mathbf{\Gamma}_m \mathbf{\Xi}_{k-m}$ and $(\mathbf{B} \cdot \mathbf{\Xi})_j = \sum_{m=0}^{\min(L-1,j)} \mathbf{B}_m \mathbf{\Xi}_{j-m}$ represent the convolutions of the sequences $\mathbf{\Gamma}_k$ and $\mathbf{B}_j$ with the Wold-decomposition $\mathbf{\Xi}_m$ of $\mathbf{x}_t$. For a given $(\mathbf{B} \cdot \mathbf{\Xi})_j$, the original coefficients $\mathbf{B}_k$ can be derived through deconvolution:

$$\mathbf{B}_k = \begin{cases} (\mathbf{B} \cdot \mathbf{\Xi})_0 \mathbf{\Xi}_0^{-1} & , k = 0 \\ \left((\mathbf{B} \cdot \mathbf{\Xi})_k - \sum_{m=0}^{k-1} \mathbf{B}_m \mathbf{\Xi}_{j-m} \right) \mathbf{\Xi}_0^{-1} & , k = 1, ..., L-1 \end{cases} \tag{6}$$

In our parametrization, $\mathbf{\Xi}_0^{-1} = \mathbf{I}$ simplifies this expression. Denote the elements of the matrices as follows:

$$(\mathbf{\Gamma} \cdot \mathbf{\Xi})_{lmk} \quad \text{and} \quad (\mathbf{B} \cdot \mathbf{\Xi})_{lmj}.$$

Assuming $\delta \geq 0$ (for forecasting and nowcasting—similar but more complex expressions would arise for $\delta < 0$ (backcasting), which are omitted here), we also assume that L is sufficiently large to render $(\mathbf{\Gamma} \cdot \mathbf{\Xi})_{\delta+k} \approx 0$ and $(\mathbf{B} \cdot \mathbf{\Xi})_j \approx 0$ for $k, j \geq L$ negligible in terms of their contributions to the variances of $\hat{\mathbf{z}}_{t\delta}$ and $\mathbf{y}_t$. We define:

$$(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{ij\delta} = ((\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ij\delta}, (\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ij\delta+1}..., (\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ij\delta+L-1})'$$

$$(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta} = \left((\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i1\delta}, (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i2\delta}, ..., (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{in\delta} \right)'$$

$$(\mathbf{b} \cdot \boldsymbol{\xi})_{ij} = ((\mathbf{B} \cdot \mathbf{\Xi})_{ij0}, (\mathbf{B} \cdot \mathbf{\Xi})_{ij1}, ..., (\mathbf{B} \cdot \mathbf{\Xi})_{ijL-1})'$$

$$(\mathbf{b} \cdot \boldsymbol{\xi})_i = \left((\mathbf{b} \cdot \boldsymbol{\xi})'_{i1}, (\mathbf{b} \cdot \boldsymbol{\xi})'_{i2}, ..., (\mathbf{b} \cdot \boldsymbol{\xi})'_{in} \right)'$$

It is important to note that if $\mathbf{\Xi}_k = \mathbf{0}$, $k > 0$, resulting in $\mathbf{x}_t = \boldsymbol{\epsilon}_t$, then we have $(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta} = \boldsymbol{\gamma}_{i \cdot \delta}$ and $(\mathbf{b} \cdot \boldsymbol{\xi})_i = \mathbf{b}_i$, as defined in the preceding section. The generalized M-SSA criterion can then be expressed as:

$$\max_{(\mathbf{b} \cdot \boldsymbol{\xi})_i} \frac{(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i\delta} \tilde{\mathbf{I}} (\mathbf{b} \cdot \boldsymbol{\xi})_i}{\sqrt{(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i\delta} \tilde{\mathbf{I}} (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta}} \sqrt{(\mathbf{b} \cdot \boldsymbol{\xi})'_i \tilde{\mathbf{I}} (\mathbf{b} \cdot \boldsymbol{\xi})_i}} \tag{7}$$

$$(\mathbf{b} \cdot \boldsymbol{\xi})'_i \tilde{\mathbf{M}} (\mathbf{b} \cdot \boldsymbol{\xi})_i = \rho_i$$

$$(\mathbf{b} \cdot \boldsymbol{\xi})'_i \tilde{\mathbf{I}} (\mathbf{b} \cdot \boldsymbol{\xi})_i = 1.$$

This formulation can be solved for $(\mathbf{b} \cdot \boldsymbol{\xi})_i$, as detailed in Theorem (1) in the next section. The M-SSA coefficients $\mathbf{b}_i$ can subsequently be determined through deconvolution, as indicated in Equation (6). Wildi (2025) posits an extension of SSA to non-stationary integrated processes, wherein the HT constraint pertains to stationary differences of the predictor. While a corresponding extension to M-SSA is feasible based on findings by McElroy and Wildi (2020), the present discussion is confined to stationary processes, emphasizing the primary topic of interest while deliberately omitting a comprehensive and potentially distracting examination of cointegration in the non-stationary case, which falls outside the current scope.

3 Solution

Wildi (2025) presents a formal framework for the univariate SSA, which we extend here to the multivariate M-SSA. For the purposes of clarity and simplicity, we assume that the process $\mathbf{x}_t = \boldsymbol{\epsilon}_t$

is WN, acknowledging that autocorrelated processes may be accommodated through the extensions proposed in the preceding section.

We consider the orthonormal Fourier eigenvectors defined as $\mathbf{v}_j := \left(\sin(k\omega_j) / \sqrt{\sum_{k=1}^{L} sin(k\omega_j)^2} \right)_{k=1,...,L}$ of the matrix $\mathbf{M}$, associated with the eigenvalues $\lambda_j = \cos(\omega_j)$ calculated at discrete Fourier frequencies $\omega_j = j\pi/(L+1)$, $j = 1, ..., L$, as detailed by Anderson (1975). We denote the solutions to the equation $\frac{\partial \rho(y_i, y_i, 1)}{\partial \mathbf{b}_i} = \mathbf{0}$ as stationary points of the lag-one ACF $\rho(y_i, y_i, 1)$.

Proposition 1. *Under the assumptions of white noise and a full-rank covariance matrix $\mathbf{\Sigma}$ as established in the previous section, the vector $\mathbf{b}_i$ constitutes a stationary point of the lag-one autocorrelation $\rho(y_i, y_i, 1)$ if and only if $\mathbf{b}_{ij} \propto \mathbf{v}_k$ for $j = 1, ..., n$, where $\mathbf{v}_k$ is an eigenvector of $\mathbf{M}$. In this scenario, the lag-one autocorrelation is given by $\rho(y_i, y_i, 1) = \lambda_k$, where λ_k corresponds to the associated eigenvalue. The extremal values of $\rho(y_i, y_i, 1)$ are obtained as $\rho_{min}(L) = \min_k \lambda_k = \cos(\pi L/(L+1)) = -\cos(\pi/(L+1))$ and $\rho_{max}(L) = \max_k \lambda_k = \cos(\pi/(L+1))$.*

Proof: For the sake of clarity, we assume that the vector $\mathbf{b}_i$ is constrained to the unit-ellipse defined by $\mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i = 1$, which implies that the autocorrelation $\rho(y_i, y_i, 1) = \mathbf{b}_i' \tilde{\mathbf{M}} \mathbf{b}_i$. To find a stationary point of $\rho(y_i, y_i, 1)$, we set the derivative of the Lagrangian $\mathfrak{L} = \mathbf{b}_i' \tilde{\mathbf{M}} \mathbf{b}_i - \lambda(\mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i - 1)$ to zero, leading to the equation $\left(\tilde{\mathbf{M}} - \lambda \tilde{\mathbf{I}} \right) \mathbf{b}_i = \mathbf{0}$. Given that $\mathbf{\Sigma}$ is assumed to be of full rank, we can manipulate the previous expression by multiplying it with $\mathbf{\Sigma}^{-1} \otimes \mathbf{I} L \times L = \tilde{\mathbf{I}}^{-1}$, yielding $(\mathbf{I}_{n \times n} \otimes \mathbf{M} - \lambda \mathbf{I}_{nL \times nL}) \mathbf{b}_i = \mathbf{0}$. This can be rewritten as $\mathbf{M} \mathbf{b}_{ij} - \lambda \mathbf{b}_{ij} = \mathbf{0}$ for $j = 1, ..., n$. Therefore, $\lambda = \lambda_k$ must be an eigenvalue of $\mathbf{M}$ for some $k \in 1, ..., L$. Furthermore, since the eigenvalues of $\mathbf{M}$ are distinct, the corresponding eigenvector $\mathbf{v}_k$ is uniquely determined, leading to the conclusion that $\mathbf{b}_{ij} \propto \mathbf{v}_k$, $j = 1, ..., n$. Additionally, we have

$$\rho(y_{ij}, y_{ij}, 1) = \frac{\mathbf{b}_{ij}' \mathbf{M} \mathbf{b}_{ij}}{\mathbf{b}_{ij}' \mathbf{b}_{ij}} = \lambda \frac{\mathbf{b}_{ij}' \mathbf{b}_{ij}}{\mathbf{b}_{ij}' \mathbf{b}_{ij}} = \lambda_k.$$

Since $\mathbf{b}_{ij} \propto \mathbf{v}_k$, we can express y_{it} as follows

$$y_{it} = \mathbf{b}_i' \boldsymbol{\epsilon}_{\cdot t} = \sum_{j=1}^{n} \mathbf{b}_{ij}' \boldsymbol{\epsilon}_{jt} = \mathbf{v}_k' \tilde{\boldsymbol{\epsilon}}_t,$$

where $\tilde{\boldsymbol{\epsilon}}_t = \sum_{j=1}^{n} a_j \boldsymbol{\epsilon}_{jt}$ is a linear combination of the components $\boldsymbol{\epsilon}_{jt}$. Given that $\tilde{\boldsymbol{\epsilon}}_t$ is also WN, we conclude that $\rho(y_i, y_i, 1) = \rho(y_{ij}, y_{ij}, 1) = \lambda_k$, as asserted. Lastly, since the unit-ellipse is devoid of boundary-points, the extremal values $\rho_{min}(L)$ and $\rho_{max}(L)$ of the lag-one ACF of y_{it} must correspond to stationary points. THus we have $\rho_{min}(L) = \min_k \lambda_k$ and $\rho_{max}(L) = \max_k \lambda_k$, which cooncludes the proof of the proposition. $\square$

We will now derive the spectral decomposition of the MSE predictor $\boldsymbol{\gamma}_{i \cdot \delta}$. Let $\tilde{\lambda}_m$, $m = 1, ..., nL$, denote the eigenvalues of the matrix $\tilde{\mathbf{M}}$, with corresponding eigenvectors $\tilde{\mathbf{v}}_m$. Specifically, we have $\tilde{\lambda}_{(k-1)n+j} = \lambda_k \tilde{\sigma}_j$ and $\tilde{\mathbf{v}}_{(k-1)n+j} = \mathbf{v}_{\sigma j} \otimes \mathbf{v}_k$, where λ_k, $\tilde{\sigma}_j$ are the eigenvalues and $\mathbf{v}_k$, $\mathbf{v}_{\sigma j}$ are the eigenvectors of $\mathbf{M}$ and $\mathbf{\Sigma}$, respectively, as established by Horn and Johnson (1991). We will index the eigenvalues and eigenvectors of $\tilde{\mathbf{M}}$ using the notation $\tilde{\lambda}_{kj}$ and $\tilde{\mathbf{v}}_{kj}$, where the two-dimensional index kj refers to the product $\lambda_k \tilde{\sigma}_j$. It is assumed that the eigenvalues λ_k are ordered in decreasing magnitude, from largest to smallest. Furthermore, we assume that the eigenvectors $\tilde{\mathbf{v}}_{kj}$ are normalized to form an orthonormal basis of $\mathbb{R}^{nL}$. The eigenvalues of $\tilde{\mathbf{I}}$ are $\tilde{\sigma}_j$, each with multiplicity L, for $j = 1, ..., n$, and the eigenvectors $\tilde{\mathbf{v}}_{kj}$ of $\tilde{\mathbf{M}}$ are also eigenvectors of $\tilde{\mathbf{I}}$. By organizing the orthonormal basis $\tilde{\mathbf{v}}_{kj}$ into the $nL \times nL$-dimensional matrix $\tilde{\mathbf{V}}$, we can express the MSE predictor $\boldsymbol{\gamma}_{i \cdot \delta}$, for $i = 1, ..., n$, in the form of the spectral decomposition:

$$\boldsymbol{\gamma}_{i \cdot \delta} = \tilde{\mathbf{V}} \mathbf{w}_i = \sum_{k=1}^{L} \sum_{j=1}^{n} w_{ikj} \tilde{\mathbf{v}}_{kj}, \tag{8}$$

where $\mathbf{w}_i = (w_{ikj})_{k=1,...,L,j=1,...,n}$ represents the spectral weights. The predictor $\boldsymbol{\gamma}_{i\cdot\delta}$ is said to possess *complete spectral support* if the condition $\sum_{j=1}^n |w_{ikj}| \neq 0$ holds for $k = 1,...,L$. This definition generalizes the concept proposed by Wildi (2024) in the univariate context, where $n = 1$ and it is required that $w_{ikj} = w_k \neq 0$ for $1 \leq k \leq L$.

Proposition 2. *Consider the M-SSA criterion defined in Equation* (5), *under the assumptions of WN and a full-rank covariance matrix* $\boldsymbol{\Sigma}$, *with* $L \geq 2$. *The following assertions hold:*

- *If* $|\rho_i| > \rho_{max}(L)$ *(exterior point) for the HT constraint of the* i-th *target, the corresponding optimization problem does not yield a solution.*

- *If* $\rho_i = \lambda_1 = \rho_{max}(L)$ *(first boundary case) and* $\sum_{j=1}^n |w_{i1j}| \neq 0$, *then the M-SSA predictor is given by*

$$\mathbf{b}_i = \frac{\sqrt{l}}{\sqrt{\sum_{j=1}^n \tilde{\sigma}_j w_{i1j}^2}} \sum_{j=1}^n w_{i1j} \tilde{\mathbf{v}}_{1j} \neq \mathbf{0},$$

  where w_{i1j} *and* $\tilde{\mathbf{v}}_{1j}$ *represent the spectral weights and eigenvectors, as specified in Equation* (8).

- *Similarly, if* $\rho_i = \lambda_L = -\rho_{max}(L)$ *(second boundary case) and* $\sum_{j=1}^n |w_{iLj}| \neq 0$, *then the M-SSA solution is expressed as*

$$\mathbf{b}_i = \frac{\sqrt{l}}{\sqrt{\sum_{j=1}^n \tilde{\sigma}_j w_{iLj}^2}} \sum_{j=1}^n w_{iLj} \tilde{\mathbf{v}}_{Lj} \neq \mathbf{0}.$$

Proof: The proof of the first assertion is derived from Proposition (1). For the second assertion, we assume $i = 1$ and analyze the M-SSA Criterion (5) defined as:

$$\max_{\mathbf{b}_1} \boldsymbol{\gamma}'_{1\cdot\delta}\tilde{\mathbf{I}}\mathbf{b}_1 \tag{9}$$
$$\mathbf{b}'_1\tilde{\mathbf{M}}\mathbf{b}_1 = \rho_1 \tag{10}$$
$$\mathbf{b}'_1\tilde{\mathbf{I}}\mathbf{b}_1 = 1. \tag{11}$$

Applying spectral decomposition to $\mathbf{b}_1$ yields:

$$\mathbf{b}_1 := \sum_{k=1}^L \sum_{j=1}^n \alpha_{kj}\tilde{\mathbf{v}}_{kj}, \tag{12}$$

where the two-dimensional index corresponds to the eigenvalues $\tilde{\lambda}_{kj} = \lambda_k\tilde{\sigma}_j$ of $\tilde{\mathbf{M}}$, with λ_k ordered from largest to smallest. The variable $\alpha_{kj} = \alpha_{1kj}$ depends on $i = 1$, and we omit the complex three dimensional indexing for clarity. Since $\tilde{\mathbf{v}}_{kj}$ serves as an eigenvector of $\tilde{\mathbf{I}}$, with strictly positive eigenvalue $\tilde{\sigma}_j > 0$, the length constraint can be reformulated as follows:

$$1 = \mathbf{b}'_i\tilde{\mathbf{I}}\mathbf{b}_i = \sum_{j=1}^n \tilde{\sigma}_j \sum_{k=1}^L \alpha_{kj}^2, \tag{13}$$

which we refer to as the elliptic (length) constraint. In a similar manner, the HT constraint (assuming $l = 1$) yields the following expression:

$$\rho_1 = \mathbf{b}'_1\tilde{\mathbf{M}}\mathbf{b}_1 = \sum_{j=1}^n \tilde{\sigma}_j \sum_{k=1}^L \lambda_k\alpha_{kj}^2$$

which we designate as the hyperbolic (HT) constraint. Given that $L \geq 2$, we can select $k = 2$ and $j = 1$ to solve for α_{21} within the hyperbolic constraint:

$$\alpha_{21}^2 = \frac{\rho_1}{\lambda_2\tilde{\sigma}_1} - \sum_{(k,j)\neq(2,1)} \frac{\lambda_k\tilde{\sigma}_j}{\lambda_2\tilde{\sigma}_1}\alpha_{kj}^2, \tag{14}$$

where the notation $\sum_{(k,j)\neq(2,1)}$ indicates summation over all combinations of (k,j) excluding the single pairing $(k,j) = (2,1)$. Similarly, we can express α_{11} in the context of the elliptic constraint as follows:

$$\alpha_{11}^2 \quad = \quad \frac{1}{\tilde{\sigma}_1} - \frac{1}{\tilde{\sigma}_1} \sum_{(k,j)\neq(1,1)} \tilde{\sigma}_j \alpha_{kj}^2,$$

which can be rewritten as

$$\alpha_{11}^2 \quad = \quad \frac{1}{\tilde{\sigma}_1} - \frac{1}{\tilde{\sigma}_1} \sum_{(k,j)\notin\{(2,1),(1,1)\}} \tilde{\sigma}_j \alpha_{kj}^2 - \left(\frac{\rho_1}{\lambda_2 \tilde{\sigma}_1} - \sum_{(k,j)\neq(2,1)} \frac{\lambda_k \tilde{\sigma}_j}{\lambda_2 \tilde{\sigma}_1} \alpha_{kj}^2 \right).$$

Here, we substitute Equation (14). This formulation characterizes the intersection of elliptic and hyperbolic constraints: if this intersection is non-existent, then the SSA-problem lacks a solution. Assuming $\rho_1 = \lambda_1 = \max_k \lambda_k$ and isolating $\tilde{\sigma}_1 \alpha_{11}^2$ to the left side of the equation, we derive:

$$\tilde{\sigma}_1 \alpha_{11}^2 \quad = \quad \frac{\lambda_2 - \rho_1}{\lambda_2 - \lambda_1} - \sum_{(k,j)\notin\{(2,1),(1,1)\}} \tilde{\sigma}_j \frac{\lambda_2 - \lambda_k}{\lambda_2 - \lambda_1} \alpha_{kj}^2 \tag{15}$$

$$= \quad 1 - \sum_{(k,j)\notin\{(2,1),(1,1:n)\}} \tilde{\sigma}_j \frac{\lambda_2 - \lambda_k}{\lambda_2 - \lambda_1} \alpha_{kj}^2 - \sum_{j=2}^{n} \tilde{\sigma}_j \frac{\lambda_2 - \lambda_1}{\lambda_2 - \lambda_1} \alpha_{1j}^2$$

$$= \quad 1 - \sum_{(k,j)\notin\{(2,1),(1,1:n)\}} \tilde{\sigma}_j \frac{\lambda_2 - \lambda_k}{\lambda_2 - \lambda_1} \alpha_{kj}^2 - \sum_{j=2}^{n} \tilde{\sigma}_j \alpha_{1j}^2.$$

where the notation $\sum_{(k,j)\notin\{(2,1),(1,1:n)\}}$ signifies that the sum extends over all combinations of (k,j) excluding $k = 1$ and the single pairing $(k = 2, j = 1)$. We can rewrite this expression as:

$$\sum_{j=1}^{n} \tilde{\sigma}_j \alpha_{1j}^2 \quad = \quad 1 - \sum_{(k,j)\notin\{(2,1),(1,1:n)\}} \tilde{\sigma}_j \frac{\lambda_2 - \lambda_k}{\lambda_2 - \lambda_1} \alpha_{kj}^2. \tag{16}$$

Under the assumption that

$$\tilde{\sigma}_j \frac{\lambda_2 - \lambda_k}{\lambda_2 - \lambda_1} \begin{cases} = 0 & k = 2 \\ < 0 & k > 2 \end{cases}$$

it follows that $\alpha_{kj} = 0$ for all pairs (k,j) not contained within $\{(2,1),(1,1:n)\}$ since otherwise the condition $\sum_{j=1}^{n} \tilde{\sigma}_j \alpha_{1j}^2 > 1$ would contradict the elliptic (length) constraint specified in Equation (13). Consequently, we deduce:

$$\sum_{j=1}^{n} \tilde{\sigma}_j \alpha_{1j}^2 = 1 \quad \text{and} \quad \alpha_{kj} = 0, k \neq 1. \tag{17}$$

The zero constraints, including the previously unmentioned $\alpha_{21} = 0$, are attributed to the elliptic constraint. It is important to note that if $\rho_1 > \lambda_1$, then the expression

$$\frac{\lambda_2 - \rho_1}{\lambda_2 - \lambda_1} > 1$$

in Equation (15) would conflict with the elliptic constraint, thereby providing an alternative proof of the initial claim. In summary, when $\rho_1 = \lambda_1$ (the first boundary case), the intersection of the ellipse and hyperbola manisfests as a reduced-rank $(n-1)$-dimensional ellipse as delineated by Equation (17). The spectral decomposition expressed in Equation (12) simplifies to

$$\mathbf{b}_1 = \sum_{j=1}^{n} \alpha_{1j} \tilde{\mathbf{v}}_{1j}.$$

Substituting this expression into the objective function represented in equation (9) results in

$$\boldsymbol{\gamma}'_{1.\delta}\tilde{\mathbf{I}}\mathbf{b}_1 = \sum_{j=1}^{n} w_{11j}\tilde{\sigma}_j\alpha_{1j},$$

leading to the corresponding boundary (reduced-rank) M-SSA problem defined as:

$$\max_{\alpha_{1j}} \sum_{j=1}^{n} w_{11j}\tilde{\sigma}_j\alpha_{1j} \text{ subject to } \sum_{j=1}^{n} \tilde{\sigma}_j\alpha_{1j}^2 = 1.$$

We can now specify the Lagrangian function

$$\mathcal{L} := \sum_{j=1}^{n} w_{11j}\tilde{\sigma}_j\alpha_{1j} - \lambda_{\mathcal{L}}\left(\sum_{j=1}^{n} \tilde{\sigma}_j\alpha_{1j}^2 - 1\right)$$

where $\lambda_{\mathcal{L}}$ is the Lagrange multiplier. The corresponding Lagrangian equations yield

$$\tilde{\sigma}_j w_{11j} - 2\lambda_{\mathcal{L}}\tilde{\sigma}_j\alpha_{1j} = 0 \ , \ j = 1, ..., n$$

which leads to the expression

$$\alpha_{1j} = \tilde{\sigma}_j w_{11j}/(2\lambda_{\mathcal{L}}).$$

Thus we can express

$$\mathbf{b}_1 = \frac{1}{2\lambda_{\mathcal{L}}}\sum_{j=1}^{n} w_{11j}\tilde{\mathbf{v}}_{1j} \neq \mathbf{0}$$

with the proportionality term

$$\frac{1}{2\lambda_{\mathcal{L}}} := \frac{1}{\sqrt{\sum_{j=1}^{n} \tilde{\sigma}_j w_{11j}^2}}$$

ensuring compliance with the elliptic length constraint. A proof of the last assertion (second boundary case) follows a similar reasoning, by symmetry. $\square$

Remarks: The aforementioned M-SSA solutions for the boundary conditions $\rho_i = \pm\rho_{max}(L)$ extend the univariate solution $\mathbf{b}_i = \begin{cases} \sqrt{l}\text{sign}(w_1)\mathbf{v}_1 \ , & \rho_i = \rho_{max}(L) \\ \sqrt{l}\text{sign}(w_L)\mathbf{v}_L \ , & \rho_i = -\rho_{max}(L) \end{cases}$ proposed by Wildi (2024,2025). In this context, it is assumed that ϵ_t are standardized, thereby omitting the standard deviation $\sigma = 1$ from the denominator. Furthermore, from the preceding proof, we deduce that if $|\rho_i| < \rho_{max}(L)$ (an interior point), the intersection of the ellipse and hyperbola results in a subspace of dimension $Ln - 2 \geq 0$ in contrast to the boundary (reduced rank) scenario described by Equation (17), which has dimension $n - 1 \geq 0$. We will subsequently derive the corresponding M-SSA predictor for the 'interior point' case, which is of primary interest in practical applications. For the purposes of this discussion, we will assume $Ln > 2$ to ensure that the optimization problem remains non-trivial.

Theorem 1. *Assuming $Ln > 2$, we define a set of regularity conditions for the M-SSA optimization problem (5) under the stipulated assumptions of WN and full-rank covariance matrix $\boldsymbol{\Sigma}$:*

1. *$\rho_i \neq \rho_{MSE}$ (non-degenerate case).*

2. *$|\rho_i| < \rho_{max}(L)$ (interior point).*

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

Then the following results hold:

1. *Under the above regularity assumptions, the solution* $\mathbf{b}i$ *to the M-SSA Criterion (5) can be expressed in a general parametric form as follows:*

$$\mathbf{b}_i = D_i \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta} = D_i \sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)} \left(\sum_{j=1}^{n} w_{ikj} \tilde{\mathbf{v}}_{kj} \right), \tag{18}$$

   where $D_i \neq 0$, $\nu_i \in \mathbb{R} - \{2\lambda_k | k = 1, ..., L\}$ *and* $\tilde{\boldsymbol{\nu}}_i := 2\tilde{\mathbf{M}} - \nu_i \tilde{\mathbf{I}}$ *is an invertible* $nL \times nL$ *matrix. Additionally,* $D(\nu_i, l)$ *is contingent on* ν_i *and the length constraint with its sign determined by the requirement for a positive objective function, specifically* $\boldsymbol{\gamma}'_{i \cdot \delta} \tilde{\mathbf{I}} \mathbf{b}_i > 0$.

2. *The lag-one acf* $\rho_i(\nu_i)$, *derived from* $\mathbf{b}_i = \mathbf{b}_i(\nu_i)$ *in Equation (18), is given by:*

$$\rho_i(\nu_i) := \frac{\mathbf{b}_i(\nu_i)' \tilde{\mathbf{M}} \mathbf{b}_i(\nu_i)}{\mathbf{b}_i(\nu_i)' \tilde{\mathbf{I}} \mathbf{b}_i(\nu_i)} = \frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 \right)}{\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 \right)} \tag{19}$$

   Any ρ_i *such that* $|\rho_i| < \rho_{max}(L)$ *is permissible under the HT constraint, implying the existence of* ν_i *such that* $\rho_i(\nu_i) = \rho_i$.

3. *The derivative* $d\rho_i(\nu_i)/d\nu_i$ *is strictly negative for* $\nu_i \in \{\nu || \nu | > 2\rho_{max}(L)\}$. *Furthermore, it holds that*

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

   where $\rho_{MSE} = \lim_{|\nu_i| \to \infty} \rho_i(\nu_i)$ *represents the lag-one ACF of the MSE predictor* $y_{it, MSE}$. *Consequently, the lag-one ACF is a strictly monotonic function for* ν_i *in the region* $\{\nu || \nu | > 2\rho_{max}(L)\}$.

4. *For* $\nu_i \in \{\nu || \nu | > 2\rho_{max}(L)\}$ *the derivatives of the objective function and the lag-one ACF as a function of* ν_i *are interconnected by the following relation:*

$$-sign(\nu_i) \frac{d\rho(y_i(\nu_i), z_i, \delta))}{d\nu_i} = \frac{\sqrt{\boldsymbol{\gamma}'_{i \cdot \delta} \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta}}}{\sqrt{\boldsymbol{\gamma}'_{i \cdot \delta} \tilde{\mathbf{I}}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta}}} \frac{d\rho_i(\nu_i)}{d\nu_i} < 0, \tag{20}$$

   where $\tilde{\boldsymbol{\nu}}_i^{-2} := \left(\tilde{\boldsymbol{\nu}}_i^{-1} \right)^2$, $y_{it}(\nu_i)$ *denotes the output of the filter with weights* $\mathbf{b}_i$, *and* $\rho(y_i(\nu_i), z_i, \delta)$ *corresponds to the target correlation and objective function.*

Proof: Consider the Lagrangian defined as

$$\mathcal{L}_i := \boldsymbol{\gamma}'_{i \cdot \delta} \tilde{\mathbf{I}} \mathbf{b}_i - \lambda_{i1}(\mathbf{b}'_i \tilde{\mathbf{I}} \mathbf{b}_i - 1) - \lambda_{i2}(\mathbf{b}'_i \tilde{\mathbf{M}} \mathbf{b}_i - \rho_i),$$

where $\lambda_{i1}, \lambda_{i2}$ denote the Lagrange multipliers. Since the intersection of the ellipse (representing the length constraint) and hyperbola (representing the HT constraint) is devoid of boundary points, the solution $\mathbf{b}_i$ to the M-SSA problem must satisify the Lagrangian equations given by

$$\tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta} = \lambda_{i1}(\tilde{\mathbf{I}} + \tilde{\mathbf{I}}') \mathbf{b}_i + \lambda_{i2}(\tilde{\mathbf{M}} + \tilde{\mathbf{M}}') \mathbf{b}_i = \lambda_{i1} 2 \tilde{\mathbf{I}} \mathbf{b}_i + \lambda_{i2} 2 \tilde{\mathbf{M}} \mathbf{b}_i.$$

The second regularity assumption (non-degenerate case) implies that the HT constraint, as expressed in Equation (10), is 'active', leading to $\lambda_{i2} \neq 0$. Dividing through by λ_{i2} yields

$$D_i \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta} = \tilde{\boldsymbol{\nu}}_i \mathbf{b}_i \tag{21}$$
$$\tilde{\boldsymbol{\nu}}_i := (2\tilde{\mathbf{M}} - \nu_i \tilde{\mathbf{I}}), \tag{22}$$

where $D_i = 1/\lambda_{i2}$ and $\nu_i = -2\frac{\lambda_{i1}}{\lambda_{i2}}$. Since the M-SSA solution is derived from the $nL - 2 > 0$ dimensional intersection of the ellipse defined in Equation (11) and the HT hyperbola chracaterized in Equation (10), we can infer that the objective function is not overruled by the constraint,

implying that $|\lambda_{i2}| < \infty$. Consequently, $D_i \neq 0$ in Equation (21) holds true. The eigenvalues of $\tilde{\boldsymbol{\nu}}_i$ are given by $\tilde{\sigma}_j(2\lambda_k - \nu_i)$, $1 \leq k \leq L$, $1 \leq j \leq n$. It is important to note that if $\mathbf{b}_i$ is a solution to the M-SSA problem, then $\nu_i/2$ cannot be an eigenvalue of $\mathbf{M}$. To illustrate this, assume $\nu_i/2 = \lambda_{k_0}$ for some $k_0 \in \{1, ..., L\}$. Under this assumption, $\tilde{\boldsymbol{\nu}}_i$ would map the eigenvectors $\tilde{\mathbf{v}}_{k_0 j}$, for $j = 1, ..., n$, to zero in Equation (21). We then look at the left-hand side of the equation:

$$D_i \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta} = D_i \tilde{\mathbf{I}} \sum_{k=1}^{L} \sum_{j=1}^{n} w_{ikj} \tilde{\mathbf{v}}_{kj} = D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj} \tilde{\mathbf{v}}_{kj}.$$

If $\tilde{\mathbf{v}}_{k_0 j}$, $j = 1, ..., n$, is mapped to zero, then $D_i \tilde{\sigma}_j w_{ik_0 j} = 0$, $j = 1, ..., n$. Consequently, $w_{ik_0 j} = 0$, $j = 1, ..., n$, (because $D_i \tilde{\sigma}_j \neq 0$), thereby contradicting the last regularity assumption (completeness). Thus, we conclude that $\nu_i/2$ cannot be an eigenvalue of $\mathbf{M}$, $\nu_i \in \mathbb{R} - \{2\lambda_k | k = 1, ..., L\}$, $\tilde{\boldsymbol{\nu}}_i^{-1}$ exists and

$$\tilde{\boldsymbol{\nu}}_i^{-1} = \tilde{\mathbf{V}} \mathbf{D}_{\nu i}^{-1} \tilde{\mathbf{V}}'.$$

The diagonal matrix $\mathbf{D}_{\nu i}^{-1}$ is characterized by well-defined entries of the form $1/(\tilde{\sigma}_j(2\lambda_k - \nu_i))$. We can proceed to solve Equation (21) for $\mathbf{b}_i$, yielding the following expression:

$$
\begin{aligned}
\mathbf{b}_i &= D_i \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i \cdot \delta} & (23) \\
&= D_i \tilde{\mathbf{V}} \mathbf{D}_{\nu i}^{-1} \tilde{\mathbf{V}}' \tilde{\mathbf{I}} \tilde{\mathbf{V}} \mathbf{w}_i \\
&= D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \frac{\tilde{\sigma}_j w_{ikj}}{\tilde{\sigma}_j(2\lambda_k - \nu_i)} \tilde{\mathbf{v}}_{kj} \\
&= D_i \sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)} \left(\sum_{j=1}^{n} w_{ikj} \tilde{\mathbf{v}}_{kj} \right). & (24)
\end{aligned}
$$

The proof of the second assertion id derived from Equation (24) and relies on the orthonormality of $\tilde{\mathbf{v}}_{kj}$:

$$
\begin{aligned}
\rho_i(\nu_i) &= \frac{\mathbf{b}_i(\nu_i)' \tilde{\mathbf{M}} \mathbf{b}_i(\nu_i)}{\mathbf{b}_i(\nu_i)' \tilde{\mathbf{I}} \mathbf{b}_i(\nu_i)} & (25) \\
&= \frac{\left(D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \frac{w_{ikj}}{2\lambda_k - \nu_i} \tilde{\mathbf{v}}_{kj} \right)' \tilde{\mathbf{M}} \left(D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \frac{w_{ikj}}{2\lambda_k - \nu_i} \tilde{\mathbf{v}}_{kj} \right)}{\left(D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \frac{w_{ikj}}{2\lambda_k - \nu_i} \tilde{\mathbf{v}}_{kj} \right)' \tilde{\mathbf{I}} \left(D_i \sum_{k=1}^{L} \sum_{j=1}^{n} \frac{w_{ikj}}{2\lambda_k - \nu_i} \tilde{\mathbf{v}}_{kj} \right)} \\
&= \frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 \right)}{\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 \right)}. & (26)
\end{aligned}
$$

It is important to note that the lag-one autocorrelation $\rho_i(\nu_i)$ of y_{it} is influenced by $\boldsymbol{\Sigma}$. Consequently, selecting $\nu_i = \nu_{0i}$ such that $\rho_i(\nu_{0i}) = \rho_i$ (in accordance with the HT constraint) indicates that the SSA solution $\mathbf{b}i(\nu0i)$ generally depends on Σ as well. As ν_i approaches $2\lambda_k$, we find that $\rho_i(\nu_i)$ converges to λ_k, for $1 \leq k \leq L$, given the assumption that $\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 > 0$ for $k = 1, ..., L$. Thus, $\rho_i(\nu_i)$ can attain the boundary values $\pm \rho_{max}(L)$. Therefore, the continuity of $\rho_i(\nu_i)$ along with the intermediate value theorem implies that any value ρ_i satisfying $|\rho_i| < \rho_{max}(L)$ is permissible in the HT constraint, as asserted.

We now proceed to prove Assertion (2). For this purpose, we define $\tilde{w}_{ik}^2 := \sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2$ in

Equation (26), resulting in the expression

$$\frac{\partial}{\partial \nu_i} \rho\Big(y_i(\nu_i), y_i(\nu_i), 1\Big) = \frac{\partial}{\partial \nu_i} \frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2}{\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2}$$

$$= 2\frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^3} \tilde{w}_{ik}^2}{\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2} - 2\frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2 \sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^3} \tilde{w}_{ik}^2}{\left(\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2\right)^2}$$

$$= 2\frac{\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^3} \tilde{w}_{ik}^2 \sum_{j=1}^{L} \frac{1}{(2\lambda_j - \nu_i)^2} \tilde{w}_{ij}^2 - \sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2 \sum_{j=1}^{L} \frac{1}{(2\lambda_j - \nu_i)^3} \tilde{w}_{ij}^2}{\left(\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2\right)^2}$$

$$= 2\left(\sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^3} \tilde{w}_{ik}^2 \sum_{j=1}^{L} \frac{1}{(2\lambda_j - \nu_i)^2} \tilde{w}_{ij}^2 - \sum_{k=1}^{L} \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2 \sum_{j=1}^{L} \frac{1}{(2\lambda_j - \nu_i)^3} \tilde{w}_{ij}^2\right) \tag{27}$$

Utilizing the fact that $\sum_{k=1}^{L} \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2 = \mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i = 1$. For the case when $k = j$, the terms cancel in Equation (27). Therefore, we assume $k \neq j$ and consider the expression:

$$\lambda_k \frac{1}{(2\lambda_k - \nu_i)^3} \tilde{w}_{ik}^2 \frac{1}{(2\lambda_j - \nu_i)^2} \tilde{w}_{ij}^2 - \lambda_k \frac{1}{(2\lambda_k - \nu_i)^2} \tilde{w}_{ik}^2 \frac{1}{(2\lambda_j - \nu_i)^3} \tilde{w}_{ij}^2$$

$$= \lambda_k \tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^2} \frac{1}{(2\lambda_j - \nu_i)^2} \left(\frac{1}{2\lambda_k - \nu_i} - \frac{1}{2\lambda_j - \nu_i}\right)$$

$$= \lambda_k \tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^3} \frac{1}{(2\lambda_j - \nu_i)^3} (2\lambda_j - 2\lambda_k)$$

$$= \tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^3} \frac{1}{(2\lambda_j - \nu_i)^3} (2\lambda_k \lambda_j - 2\lambda_k^2)$$

$$= \tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^3} \frac{1}{(2\lambda_j - \nu_i)^3} (2\lambda_k \lambda_j - 2\lambda_j^2),$$

where we interchanged k and j in the last equality, owing to symmetry. Assuming $k \neq j$ and considering the addition of the two symmetric terms corresponding to (k, j) and (j, k), we derive the following expression:

$$\tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^3} \frac{1}{(2\lambda_j - \nu_i)^3} 2 \left(2\lambda_k \lambda_j - \lambda_k^2 - \lambda_j^2\right)$$

$$= -2\tilde{w}_{ik}^2 \tilde{w}_{ij}^2 \frac{1}{(2\lambda_k - \nu_i)^3} \frac{1}{(2\lambda_j - \nu_i)^3} (\lambda_k - \lambda_j)^2 < 0,$$

where the strict inequality is justified by the conditions $|\nu_i| > 2\rho_{max}(L) \geq 2|\lambda_k|$, $\lambda_k \neq \lambda_j$, and the assumption $\tilde{w}_{ik}^2 = \sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2 > 0$. Consequently, We infer that

$$\frac{\partial}{\partial \nu_i} \rho\Big(y_i(\nu_i), y_i(\nu_i), 1\Big) < 0.$$

Furthermore, from Equation (26) we can deduce the limit

$$\lim_{|\nu_i| \to \infty} \rho(\nu_i) = \frac{\sum_{k=1}^{L} \lambda_k \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2\right)}{\sum_{k=1}^{L} \left(\sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2\right)} = \frac{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{M}} \boldsymbol{\gamma}_{i\cdot\delta}}{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}} = \rho_{MSE},$$

which represents the lag-one ACF of the MSE predictor. In conjunction with the identified strictly negative derivative, we conclude that $\rho\Big(y_i(\nu_i), y_i(\nu_i), 1\Big)$ is a strictly monotonic function of ν_i

14

within the domain $\{\nu \mid |\nu| > 2\rho_{max}(L)\}$, as asserted.

For a proof of the last Assertion (4) we analyze the expession

$$
\begin{aligned}
\rho(y_i(\nu_i), z_i, \delta) &= \frac{\mathbf{b}_i' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{\sqrt{\mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}} = D_i \frac{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{\sqrt{D_i^2 \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1}{}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}} \\
&= D_i \frac{\left(\sum_{k=1}^{L} \frac{1}{2\lambda_k - \nu_i} \sum_{j=1}^{n} w_{ikj} \tilde{\mathbf{v}}_{kj}'\right)\left(\sum_{k=1}^{L} \sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj} \tilde{\mathbf{v}}_{kj}\right)}{\sqrt{D_i^2 \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1}{}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}} \\
&= \text{sign}(D_i) \frac{\sum_{k=1}^{L} \frac{1}{2\lambda_k - \nu_i} \sum_{j=1}^{n} \tilde{\sigma}_j w_{ikj}^2}{\sqrt{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1}{}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}}.
\end{aligned}
$$

For the case where $\nu < -2\rho_{max}(L)$, the quotient is strictly positive. Consequently, the positivity of the objective function for the M-SSA solution implies that $\text{sign}(D_i) = -\text{sign}(\nu) = 1$. Conversely, for $\nu > 2\rho_{max}(L)$ the quotient becomes strictly negative, leading to $\text{sign}(D_i) = -\text{sign}(\nu) = -1$. Assuming $\nu < -2\rho_{max}(L)$ we deduce:

$$
\begin{aligned}
\frac{d\rho\big(y_i(\nu_i), z_i, \delta\big)}{d\nu_i} &= \frac{d}{d\nu_i}\left(\frac{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{\sqrt{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}^2 \boldsymbol{\gamma}_{i\cdot\delta}}}\right) \\
&= \frac{1}{\left(\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}\right)^{3/2} \sqrt{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}^2 \boldsymbol{\gamma}_{i\cdot\delta}}} \left\{\left(\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}\right)^2 - \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-3} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}\right\} \\
&= -\text{sign}(\nu) \frac{\sqrt{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}}{\sqrt{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}^2 \boldsymbol{\gamma}_{i\cdot\delta}}} \frac{d\rho_i(\nu_i)}{d\nu_i} < 0,
\end{aligned}
$$

where we inserted $-\text{sign}(\nu) = 1$. The last expression is obtained by recognizing that

$$
\begin{aligned}
\frac{d\rho_i(\nu_i)}{d\nu_i} &= \frac{d}{d\nu_i}\left(\frac{\mathbf{b}_i' \tilde{\mathbf{M}} \mathbf{b}_i}{\mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i}\right) = \frac{d}{d\nu_i}\left(\frac{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1}{}' \tilde{\mathbf{M}} \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1}{}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}\right) = \frac{d}{d\nu_i}\left(\frac{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\mathbf{M}} \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}\right) \\
&= \frac{2\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\mathbf{M}} \tilde{\boldsymbol{\nu}}_i^{-3} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \cdot \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} - 2\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\mathbf{M}} \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \cdot \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-3} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{(\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta})^2} \qquad (28) \\
&= \frac{(\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta})^2 - \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} \cdot \boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-3} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}}{(\boldsymbol{\gamma}_{i\cdot\delta}' \tilde{\mathbf{I}}' \tilde{\boldsymbol{\nu}}_i^{-2} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta})^2}, \qquad (29)
\end{aligned}
$$

where $\tilde{\boldsymbol{\nu}}_i^{-k} := (\tilde{\boldsymbol{\nu}}_i^{-1})^k$, $(\tilde{\boldsymbol{\nu}}_i^{-1})' = \tilde{\boldsymbol{\nu}}_i^{-1}$ (by symmetry). The commutativity of the matrix product utilized in deriving the third equality arises from the simultaneous diagonalization of both $\tilde{\mathbf{M}}$ and $\tilde{\boldsymbol{\nu}}_i^{-1}$, which share the same eigenvectors. Additionally, standard rules for matrix differentiation were employed in the derivation of Equation (28)[1]. Finally, the expression $2\tilde{\mathbf{M}} \tilde{\boldsymbol{\nu}}_i^{-k} = \tilde{\boldsymbol{\nu}}_i^{-k+1} + \nu \tilde{\boldsymbol{\nu}}_i^{-k}$ was incorporated into the numerator of Equation (28), leading to the final equation after simplification. The aforementioned proof is also applicable in the scenario where $\nu > 2\rho_{\max}(L)$, albeit with a change in sign, such that $\text{sign}(D_i) = -\text{sign}(\nu) = -1$, as was to be demonstrated. $\qquad \square$

Remarks

As $|\nu_i| \to \infty$ and $D_i = -\nu_i$, it follows that $-\tilde{\boldsymbol{\nu}}_i / \nu_i \to \tilde{\mathbf{I}}$. Consequently, we have

$$
\mathbf{b}_i \to -\frac{D_i}{\nu_i} \tilde{\mathbf{I}}^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta} = \boldsymbol{\gamma}_{i\cdot\delta}.
$$

---

[1] $\frac{d(\tilde{\boldsymbol{\nu}}_i^{-1})}{d\nu_i} = \tilde{\boldsymbol{\nu}}_i^{-2}$ and $\frac{d(\tilde{\boldsymbol{\nu}}_i^{-2})}{d\nu_i} = 2\tilde{\boldsymbol{\nu}}_i^{-3}$. The first equation follows from the general differentiation rule $\frac{d(\tilde{\boldsymbol{\nu}}_i^{-1})}{d\nu_i} = -\tilde{\boldsymbol{\nu}}_i^{-1} \frac{d\tilde{\boldsymbol{\nu}}_i}{d\nu_i} \tilde{\boldsymbol{\nu}}_i^{-1}$, noting that $\frac{d\tilde{\boldsymbol{\nu}}_i}{d\nu_i} = -\tilde{\mathbf{I}}$. The second equation is derived by substituting the first equation into $\frac{d(\tilde{\boldsymbol{\nu}}_i^{-2})}{d\nu_i} = \frac{d(\tilde{\boldsymbol{\nu}}_i^{-1})}{d\nu_i} \tilde{\boldsymbol{\nu}}_i^{-1} + \tilde{\boldsymbol{\nu}}_i^{-1} \frac{d(\tilde{\boldsymbol{\nu}}_i^{-1})}{d\nu_i}$.

This indicates that the M-SSA method asymptotically replicates the MSE predictor in the limit as $|\nu_i| \to \infty$, reflecting an asymptotically degenerate case where the HT constraint may be omitted. By virtue of strict monotonicity, as stated in Assertion (3), we conclude that M-SSA can accommodate HT constraints $\rho_i < \rho_{MSE}$ when $\nu_i \in]-\infty, -2\rho_{max}(L)]$. Conversely, for $\nu_i \in [2\rho_{max}(L), \infty[$ M-SSA is capable of accommodating HT constraints $\rho_i > \rho_{MSE}$ (smoothing). Moreover, Equation (20) elucidates a trade-off between Accuracy (target correlation) and Smoothness (lag-one ACF) in the context of solving the M-SSA optimization problem. For a more comprehensive discussion, refer to Wildi (2024) and (2025). Finally, Wildi (2025) also addresses the singular case of incomplete spectral support, which may be generalized to M-SSA; however, this topic is omitted here for the sake of brevity.

Corollary 1. *Let the assumptions of theorem* (1) *be satisfied.*

1. *For each $i = 1, ..., n$, the solution $\mathbf{b}_i = (\mathbf{b}'_{i1}, \mathbf{b}'_{i2}, ..., \mathbf{b}'_{in})'$ to the M-SSA optimization problem can alternatively be derived from the n subsystems given by*

$$\mathbf{b}_{ij} = D_i \boldsymbol{\nu}_i^{-1} \boldsymbol{\gamma}_{ij\delta}, \tag{30}$$

   for $j = 1, ..., n$, where $D_i \neq 0$ and $\boldsymbol{\nu}_i := 2\mathbf{M} - \nu_i \mathbf{I}$ represents an invertible $L \times L$ matrix. The subsystems are interconnected through the common parameters D_i and ν_i, while they generally differ with respect to $\boldsymbol{\gamma}_{ij\delta}$, $j = 1, ..., n$.

2. *For each $i = 1, ..., n$, the M-SSA predictor satsifies the following non-stationary and time reversible difference equation:*

$$b_{ijk+1} - \nu_i b_{ijk} + b_{ijk-1} = D_i \gamma_{ijk+\delta} , \ 0 \leq k \leq L - 1, \ j = 1, ..., n, \ i = 1, ..., n \tag{31}$$

   with implicit boundary conditions $b_{ij,-1} = b_{ijL} = 0$ to ensure stability of the solution.

Proof: The first claim can be substantiated by multiplying both sides of Equation (21) by $\tilde{\mathbf{I}}^{-1} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_{L \times L}$. This yields the following equations:

$$\begin{aligned}
D_i \boldsymbol{\gamma}_{i \cdot \delta} &= \tilde{\mathbf{I}}^{-1} \tilde{\boldsymbol{\nu}}_i \mathbf{b}_i \\
&= \left(2(\mathbf{I}_{n \times n} \otimes \mathbf{M}) - \nu_i \mathbf{I}_{nL \times nL} \right) \mathbf{b}_i.
\end{aligned}$$

A proof of this claim is facilitated by the block-diagonal structure of $\mathbf{I}_{n \times n} \otimes \mathbf{M}$. The validity of the second claim follows directly from Equation (30), achieved by multiplying both sides by $\boldsymbol{\nu}_i$. Specifically, the boundary equations $b_{ij1} - \nu_i b_{ij0} = D_i \gamma_{ij\delta}$ and $-\nu_i b_{ij(L-1)} + b_{ij(L-2)} = D_i \gamma_{ij(L-1+\delta)}$ at $k = 0$ and $k = L - 1$ substantiate the implicit boundary conditions $b_{ij,-1} = b_{ijL} = 0$.

Theorem (1) presents a one-parameter formulation of the SSA solution, and the subsequent corollary elucidates this solution by establishing a relationship between the unknown parameter ν_i and the HT constraint.

Corollary 2. *Let the assumptions of theorem* (1) *be satisfied. Then, for each $i = 1, ..., n$, the solution to the SSA-optimization problem described in Equation* (5) *is given by $s\mathbf{b}_i(\nu_{0i})$, where $\mathbf{b}_i(\nu_{0i})$ is derived from Equation* (18). *This is under the assumption of an arbitrary scaling such that $|D_i| = 1$ (with the sign of D_i determined by the requirement for a positive objective function). Here, ν_{0i} represents a solution to the non-linear HT equation $\rho(\nu_i) = \rho_i$, and s is defined as $s = \sqrt{l / \mathbf{b}_i(\nu_{0i})' \tilde{\mathbf{I}} \mathbf{b}_i(\nu_{0i})}$. If the search for an optimal ν_{0i} can be confined to the domain $\{\nu | |\nu| > 2\rho_{max}(L)\}$, then ν_{0i} is uniquely determined by ρ_i.*

A proof follows directly from Theorem (1), with the observation that the scaling factor $s = \sqrt{l / \mathbf{b}_i(\nu_{0i})' \tilde{\mathbf{I}} \mathbf{b}_i(\nu_{0i})}$ does not interfere with either the objective function or the HT constraint. This scaling can be applied subsequently, after obtaining a solution under the assumption of an

arbitrary scaling such that $|D| = 1$. Notably, the final assertion of the theorem guarantees the uniqueness of solutions within the set $\{\nu, |, |\nu| > 2\rho_{\max}(L)\}$. $\qquad\square$

Remark: The objective function $\rho(y_i, z_{i\delta}, \delta)$ and the lag-one ACF $\rho(y_i, y_i, 1)$ are generally influenced by $\boldsymbol{\Sigma}$. Interestingly, Equation (30) indicates that the one-parameter form of the solution $\mathbf{b}_i(\nu_i)$ is independent of $\boldsymbol{\Sigma}$; this is corroborated by the cancellation of $\tilde{\sigma}_j$ in Equation (24). However, the optimal solution $\mathbf{b}_i(\nu_{0i})$, which is derived from the optimal ν_{0i}, remains dependent on $\boldsymbol{\Sigma}$ because $\tilde{\sigma}_j$ (for $j = 1, \ldots, n$) does not cancel in the lag-one ACF (19), which is critical for determining ν_{i0} within the context of the nonlinear HT constraint.

The next result addresses the distribution of the M-SSA predictor.

Corollary 3. *Let all regularity assumptions of Theorem (1) be satisified, and let $\hat{\boldsymbol{\gamma}}_{i\cdot\delta}$ represent a finite-sample estimate of the MSE-predictor $\boldsymbol{\gamma}_{i\cdot\delta}$, with mean $\boldsymbol{\mu}_{\gamma_{i\delta}}$ and variance $\boldsymbol{\Sigma}_{\gamma_{i\delta}}$. Then, the mean and variance of the M-SSA predictor $\hat{\mathbf{b}}_i$ are expressed as follows:*

$$
\begin{aligned}
\boldsymbol{\mu}_{\mathbf{b}_i} &= D_i \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\mu}_{\gamma_{i\delta}} \\
\boldsymbol{\Sigma}_{\mathbf{b}_i} &= D_i^2 \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\Sigma}_{\gamma_{i\delta}} \tilde{\mathbf{I}} \tilde{\boldsymbol{\nu}}_i^{-1}
\end{aligned}
$$

If $\hat{\boldsymbol{\gamma}}_{i\cdot\delta}$ follows a Gaussian distribution, then $\hat{\mathbf{b}}_i$ will also be Gaussian distributed.

The proof follows directly from Equation (18). For a comprehensive derivation of the mean, variance, and (asymptotic) distribution of the MSE estimate, one may consult standard textbooks, such as Brockwell and Davis (1993). In conclusion, the final result in this section introduces a dual reformulation of the M-SSA optimization criterion, distinguishing its solution as the smoothest among all predictors that maintain equivalent tracking accuracy or target correlation.

Corollary 4. *Under the assumptions stipulated in Theorem (1), let $y_{it}(\nu_{0i})$ represent the M-SSA solution for the i-th target, with the assumption that $\nu_{0i} > 2\rho_{max}(L)$. Define $\rho_{\nu_{0i},\delta} := \rho(y_i(\nu_{01}), z_i, \delta)$ and consider the dual optimization problem given by*

$$
\left.
\begin{aligned}
&\max_{\mathbf{b}_i} \rho(y_i, y_i, 1) \\
&\rho(y_i, z_i, \delta) = \rho_{\nu_{0i},\delta} \\
&\mathbf{b}_i' \tilde{\mathbf{I}} \mathbf{b}_i = 1
\end{aligned}
\right\}, \tag{32}
$$

where the objective function and the HT constraint are interchanged. If the search for an optimal ν_i can be confined to the domain $\{\nu | |\nu| > 2\rho_{max}(L)\}$, then $y_{it}(\nu_{0i})$ remains the solution to the dual problem. Conversely, if $\nu_{0i} < -2\rho_{max}(L)$, then $y_{it}(\nu_{0i})$ remains the solution to the dual problem provided that minimization replaces maximization in the objective of Criterion (32).

Proof: We begin by considering the case where $\nu_{0i} > 2\rho_{max}(l)$. The Lagrangian Equation (21) does not differentiate between constraint and objective. Through appropriate rescaling of the multipliers, the problem outlined in Criterion (32) results in an identical functional form for its solution, expressed as $\mathbf{b}_i = D_i \tilde{\boldsymbol{\nu}}_i^{-1} \tilde{\mathbf{I}} \boldsymbol{\gamma}_{i\cdot\delta}$. This is possible because the regularity assumptions ensure that the multipliers are non-vanishing and of finite size. The only distinction lies in the selection of ν_i within Criterion (32), which must satisfy the condition $\rho(y_i(\nu_i), z_i, \delta) = \rho_{\nu_{0i},\delta}$. If the search space can be restricted to $\{\nu | \nu > 2\rho_{max}(L)\}$, then according to Assertion (4), the solution to the primal problem is also a solution to the dual problem, a consequence of the strict monotonicity of $\rho(y_i(\nu_i), z_i, \delta)$ with respect to ν_i. The extension to the set $\{\nu | |\nu| > 2\rho_{max}(L)\}$ is supported by Assertion (3), which affirms that $\rho(y_i(\nu_i), y_i(\nu_i), 1) = \rho_i(\nu_i) < \rho_i(\nu_{0i})$ when $\nu_i < -2\rho_{max}(L)$. A similar line of reasoning applies to the case where $\nu_{0i} < -2\rho_{max}(L)$, with the necessary adjustment that maximization in the dual Criterion (32) must be substituted with minimization, as $\rho_i(\nu_i) > \rho_i(\nu_{i0})$ if $\nu_i > 2\rho_{max}(L)$. $\qquad\square$

4 Applications

Our examples underscore three key aspects: smoothing, forecasting, and nowcasting, or real-time signal extraction. Smoothing is primarily concerned with a causal target. In contrast, traditional forecasting focuses typically on an acausal, forward-looking allpass filter. Signal extraction encompasses various acausal filtering techniques, including lowpass filters for trend extraction, bandpass filters for cycle analysis, highpass filters for separating cycles from noise, and seasonal adjustment.

4.1 Forecasting

We apply M-SSA to a stationary VAR(1) process represented by the equation:

$$
\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \mathbf{A}_1 \begin{pmatrix} z_{1t-1} \\ z_{2t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \quad , \quad \mathbf{A}_1 = \begin{pmatrix} 0.7 & 0.4 \\ -0.6 & 0.9 \end{pmatrix} \quad , \quad \mathbf{\Sigma} = \begin{pmatrix} 1.09 & -1.45 \\ -1.45 & 2.58 \end{pmatrix}.
$$

We compute one-step ahead predictors for z_{1t+1} and z_{2t+1} with $\delta = 1$ and a predictor length of $L = 100$. The HTs within the constraint are defined as $ht_1 = 3$, $ht_2 = 8$. The target $\mathbf{z}_{t+\delta}$ is characterized as an anticipative or acausal allpass with weights defined as follows:

$$
\Gamma_k = \begin{cases} \mathbf{Id} & k = -\delta = -1 \\ \mathbf{0} & \text{otherwise.} \end{cases}
$$

The VAR process allows for a convergent MA-inversion with weights $\mathbf{\Xi}_k = \mathbf{A}^k$. Accordingly, the HTs of the MSE predictor can be computed by substituting the lag-one ACFs $(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i\delta} \tilde{\mathbf{M}} (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta} / (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i\delta} \tilde{\mathbf{I}} (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta}$, $i = 1, ..., 2$, into Equation (3) to yield $ht_{1MSE} = 5.6$ and $ht_{2MSE} = 4.6$. Here, $(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta}$ represents the convolution of the target and the MA-inversion, as discussed in Section (2.4). Given that the target is a forward-shifting allpass filter, we conclude that $(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{i\delta}$ corresponds to the shifted original MA-inversion, with the lead $-\delta = -1$ omitted.

With our specifications we observe $ht_1 < ht_{1MSE}$, indicating that the M-SSA predictor must introduce additional zero crossings to 'unsmooth' the MSE benchmark for the first series. Conversely, for the second series, $ht_2 > ht_{2MSE}$ suggests that the M-SSA predictor must 'smooth' the MSE benchmark by producing fewer zero crossings. M-SSA and MSE predictors are displayed in Fig.(1), with the M-SSA predictor presented in the top panel and the MSE predictor in the bottom panel. The weights associated with the MSE predictor are defined as $\mathbf{A}_1^\delta = \mathbf{A}_1$. The unsmoothing effect is characterized by an alternating high frequency pattern exhibited by the M-SSA filter in the first series (top left panel). In contrast, the smoothing effect results in slowly monotonically decaying patterns that are observed in the second series (top right panel).
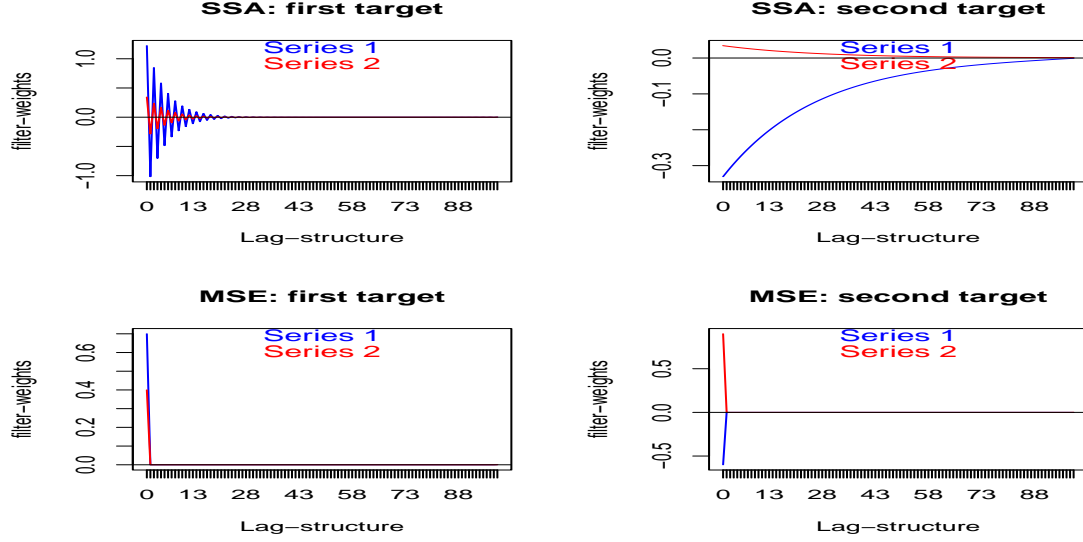
Figure 1: M-SSA (top) and MSE (bottom) predictors for the first target series $z_{1t+\delta}$ (left) and the second target series $z_{2t+\delta}$ (right) with $\delta = 1$ (one-step ahead). Predictor weights applied to first series z_{1t} (blue) and second series z_{2t} (red).

We proceed to compute sample estimates of the relevant performances measures for a lengthy sample, $N = 100000$, which ensures a near convergence of empirical to expected measures, as presented in Table (1). The final column of the table corroborates the anticipated HTs of the MSE predictor, as derived previously. The optimal values of ν_i specified in Corollary (2) are ν_1=-2.034$\in]-\infty, -2\rho_{max}(L)]$ (indicating unsmoothing) and ν_2=2.001$\in [2\rho_{max}(L), \infty[$ (indicating smoothing), as alluded to by the final remark made after Theorem (1). A comparative analysis

	Sample crit.	True crit.	Sample ht SSA	True ht SSA	Sample ht MSE
Series 1	0.91	0.91	3.02	3.00	5.61
Series 2	0.66	0.67	8.04	8.00	4.65

Table 1: Performances of SSA- and MSE-predictors: empirical and true holding-times as well as criterion values (correlations of SSA with MSE-predictors).

of the sample predictors and targets is presented in Figure (2), which demonstrates the effects of 'unsmoothing' in the first series and smoothing in the second series, as dictated by the HT constraints.

19
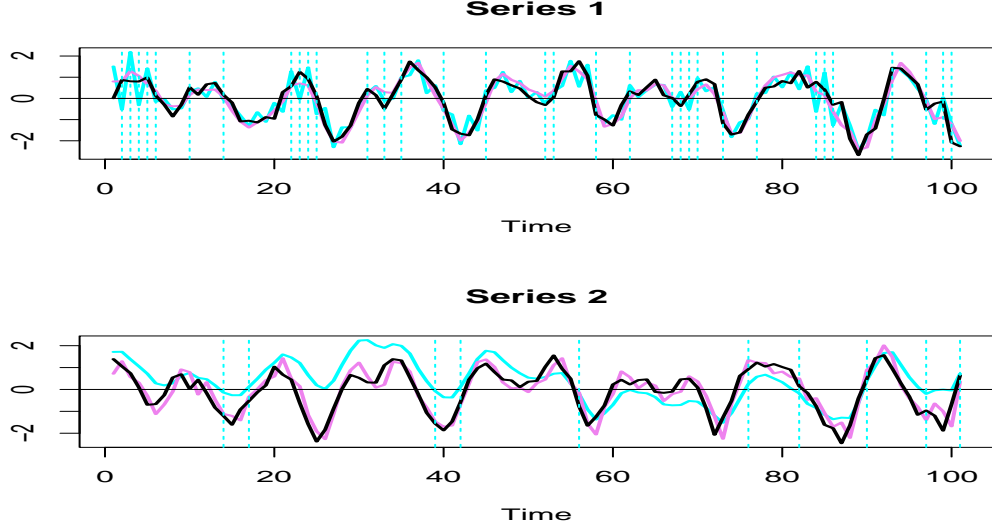
Figure 2: A comparison of filter outputs: M-SSA (cyan), MSE (violet) and target (black) for $\delta = 1$. Zero-crossings by M-SSA are marked by vertical lines.

4.2 Smoothing

If the target correlation in the previous example is computed against the *acausal* target $\mathbf{z}_{t+1}$, then M-SSA functions as a forecasting method. However, as discussed in Section (2.2), the MSE predictor can be utilized in place of the effective target without impacting the optimization results. In this context, M-SSA is then considered a smoother for the *causal* MSE predictor, with smoothing referring specifically to the application of M-SSA to causal targets. In this framework, we can examine the well-established Whittaker-Henderson (WH) graduation, expressed as follows:

$$\min_{\mathbf{b}_i} \sum_{t=L}^{T} (x_t - y_t)^2 + \lambda \sum_{t=L+D}^{T} (\Delta^D y_t)^2, \tag{33}$$

where Δ^D represents the common difference operator applied D-times (refer to Orfanidis, 2007). Notably, for $D = 2$, this formulation yields the Hoddrick-Prescott (HP) filter, as detailed by Hodrick and Prescott (1997). The equation delineates a tradeoff between conflicting objectives of data-fitting and data-smoothing terms, with the regularization parameter λ serving to balance this trade-off. An alternative tradeoff can be formalized in the (univariate) SSA-framework

$$\left. \begin{array}{c} \max_{\mathbf{b}} \rho\Big(y, x, \delta \leq 0\Big) \\ \rho(y, y, 1) = \rho_1 \end{array} \right\}, \tag{34}$$

where $z_t = x_{t+\delta}$, such that the target filter $\gamma_k = \left\{ \begin{array}{ll} 1 & k = -\delta \geq 0 \\ 0 & \text{otherwise} \end{array} \right.$ represents a causal, backward-looking (back-shifting) allpass filter, as introduced and discussed in Wildi (2024). When $\delta = 0$, the target filter becomes the identity, $z_t = x_t$. Conversely, , if L is defined as an odd integer and we select $\delta = -(L-1)/2$, this ensures symmetry in the backcast; the coefficients of the causal filter are centered at $x_{t-(L-1)/2}$, with the tails reflecting symmetrically around this central point. For the range $-(L-1)/2 \leq \delta \leq 0$, the design transitions from symmetry to asymmetry as it approaches the nowcast at $\delta = 0$. To illustrate this concept, Fig.(3) presents the SSA smoothers based on Criterion (34), scaled arbitrarily to unit length such that $\mathbf{b}'\mathbf{b} = 1$, plotted as a function

of δ. This analysis assumes WN data, with $L = 201$ and $\rho_1 = 0.9986$, which is inserted into the HT constraint of Criterion (34). Here, ρ_1 represents the lag-one ACF of the two-sided HP filter with parameter $\lambda = 14400$, which corresponds to a monthly sampling frequency, as discussed by Ravn and Uhlig (2002). The SSA smoothers replicate exactly the HT of HP, regardeless of δ. Furthermore, as indicated by Corollary (4), the symmetric SSA smoother is expected to yield superior performance compared to the two-sided HP filter with respect to target correlation, contingent upon the assumption of white noise data. This assertion will be examined in greater detail in the subsequent sections. As δ approaches zero, the smoothers converge towards $\mathbf{v}_1$ (depicted as the black line in the figure), which is the eigenvector of $\mathbf{M}$ associated with its largest eigenvalue $\lambda_1 = 0.99988 = \rho_{max}(L)$. This convergence occurs because $\rho_1 \approx \rho_{max}(L)$ is close to the upper limit of admissibility for the lag-one ACF, as established in Proposition (2).
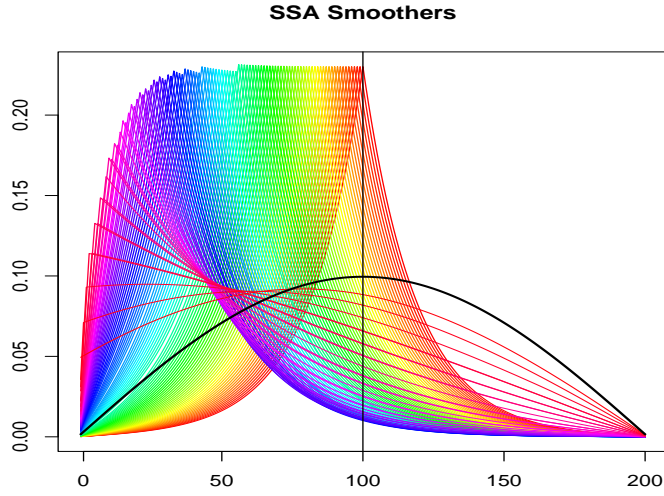
SSA Smoothers

Figure 3: Eigenvector $\mathbf{v}_1$ of $\mathbf{M}$ (black) and SSA smoothers scaled to unit length as a function of δ with values ranging from $\delta = -(L-1)/2 = -100$ (symmetric filter) to $\delta = 0$ (asymmetric nowcast), assuming the data to be WN and imposing the HT (or lag-one ACF) of the two-sided HP(14400) filter.

The proposed optimization problem can be regarded as a novel smoothing algorithm, distinct from the WH and HP methods, in that the regularization term is framed as a HT constraint. It is important to note that we have not explicitly specified a length constraint in Criterion (34); however, such a constraint is implicitly assumed for identification purposes. We can now draw a comparison between our approach and the HP method, which is effectively WH with $D = 2$. In the context of stationary data, an increase in λ typically results in an extended holding time for the HP(λ) filter. In contrast, Criterion (34) is more effective in regulating this specific characteristic. For illustration, Figure (4) presents a truncated symmetric causal HP(14400) filter alongside two distinct symmetric SSA smoothers, with the assumption that $\delta = -\frac{(L-1)}{2}$. Additionally, Table (2) provides a comparative analysis of the performance of all three designs, assuming the data follows a WN sequence. For a given identical holding time, the SSA1 design (represented by the blue line in the figure) demonstrates superior performance in target correlation compared to HP. In the dual reformulation of Corollary (4), the SSA2 design (depicted by the violet line in the figure) exhibits a maximal HT, surpassing the HP filter, given an equivalent target correlation. However, when evaluating noise suppression based on mean-square second-order differences, HP outperforms both SSA smoothers due to its inherent design. The observed variations across the reported perfor-

	HP	SSA1	SSA2
Holding times	59.548	59.548	75.000
Target correlations	0.205	0.228	0.205
RMS second-order differences	0.005	0.024	0.017

Table 2: HP vs. two different SSA smoothers. SSA1 replicates the holding time of HP and SSA2 replicates its target correlation. Root mean-square second-order differences in the last row refer to standardized white noise data.

mance metrics are significant enough to necessitate a well-informed decision regarding whether to minimize the rate of zero-crossings using SSA or to minimize curvature using WH. Notably, the exclusivity of this choice can be alleviated by integrating both methodologies, specifically by replacing the identity by the HP trend in the objective function of Criterion (34). This approach, termed 'benchmark customization' in Wildi (2024), is detailed in the following section on signal extraction. In particular, the resultant hybrid design facilitates the optimal tracking of a specified benchmark predictor (or smoother) by the SSA, while simultaneously conforming to a predetermined HT. In conclusion regarding the topic of SSA smoothing, we note that the shape of the smoothers in our examples is determined by the assumption of WN. For autocorrelated data, the extension discussed in Section (2.4) is applicable, resulting in corresponding effects on the shape of the smoothers.
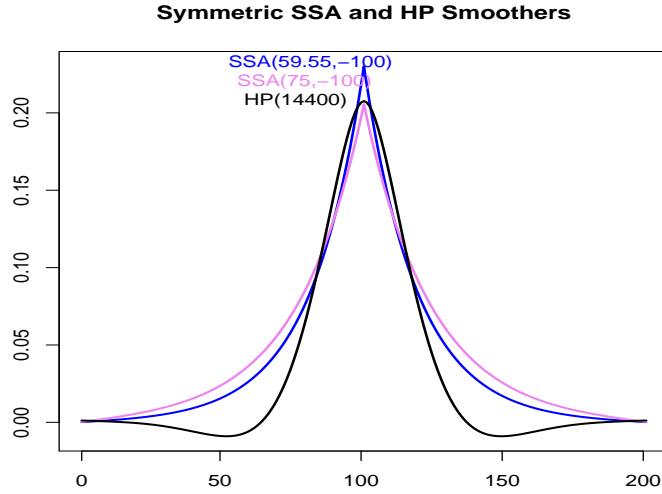
Symmetric SSA and HP Smoothers

Figure 4: Coefficients of symmetric causal SSA and HP smoothers of length 201, arbitrarily scaled to unit length: the first SSA design (blue line) replicates the holding time of HP, the second SSA design (violet line) replicates the tracking-ability or target correlation of HP.

We can now readily extend the aforementioned SSA smoothing concept to accommodate dependency within a multivariate design. For illustrative purposes, consider a straightforward three-dimensional Vector Autoregressive (VAR) process of order one, denoted as VAR(1),

$$\mathbf{x}_t = \begin{pmatrix} 0.7 & 0.4 & -0.2 \\ -0.6 & 0.9 & 0.3 \\ 0.5 & 0.2 & -0.3 \end{pmatrix} \mathbf{x}_{t-1} + \boldsymbol{\epsilon}_t \ , \ \boldsymbol{\Sigma} = \begin{pmatrix} 3.17 & 0.77 & -0.5 \\ 0.77 & 0.69 & 0 \\ -0.5 & 0 & 1.7 \end{pmatrix}$$

with Wold-decomposition $\boldsymbol{\Xi}_k = \mathbf{A}^k$, assuming a causal allpass target with $\delta = 0$, which corre-

sponds to the proper identity. We select $L = 51$ and set the HTs $ht_1 = 8$, $ht_2 = 6$ and $ht_3 = 10$ within the HT constraints for the three M-SSA smoothers y_{1t}, y_{2t} and y_{3t}, as illustrated in Fig.(5). The plots indicate that the second series x_{2t} (in red) serves as a significant explanatory variable for all three target series (after suitable adjustment of scales). Fig.(6) provides insight into this phenomenon, as the peaks of the CCF shown in the left panel, along with the original time series in the right panel, indicate that x_{2t} is left-shifted or leading. This finding can be interpreted as supporting the importance of a leading indicator within a multivariate nowcasting framework. Table (3) presents expected and sample performance measures, the latter in parentheses, based on an extensive sample size $N = 100000$, which corroborates the posited constraints. Additionally, the HTs derived from the original (unfiltered) data, detailed in the last column, identify the second series as exhibiting the smoothest characteristics among the three. Consequently, the corresponding filter weights depicted in Fig.(5) (mid panel, red) are permitted to decay more rapidly towards zero, thereby approaching the identity, as the imposed HT does not significantly deviate from the original data. The dual characteristics of x_{2t} being both a leading and smoother variable further reinforce its significance as an explanatory factor in the context of endpoint smoothing and forecasting.
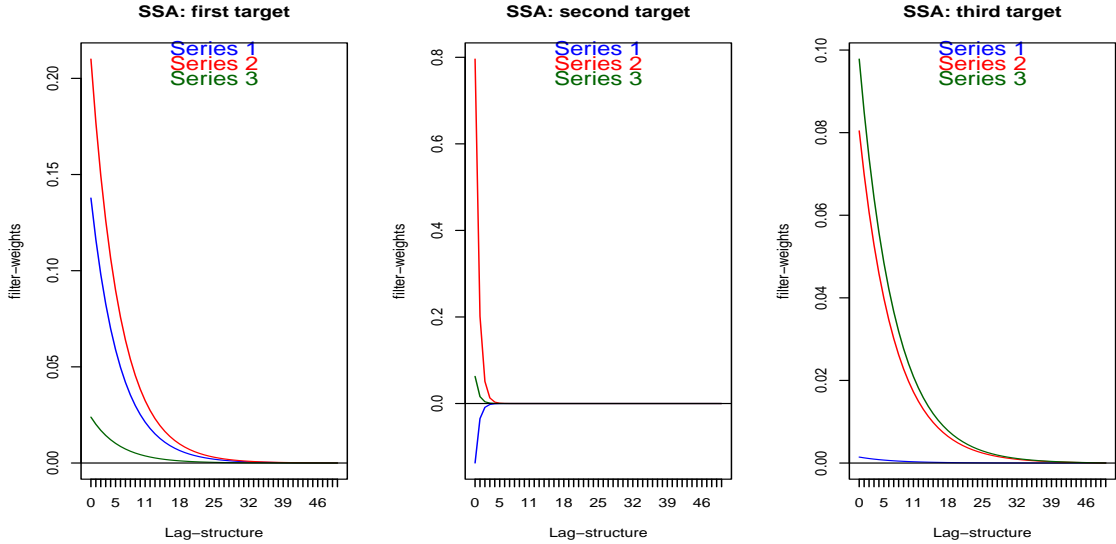

Figure 5: M-SSA smoother: first target series (left), second target (middle) and third target (right). Filters applied to first series (blue), second series (red) and third series (green).
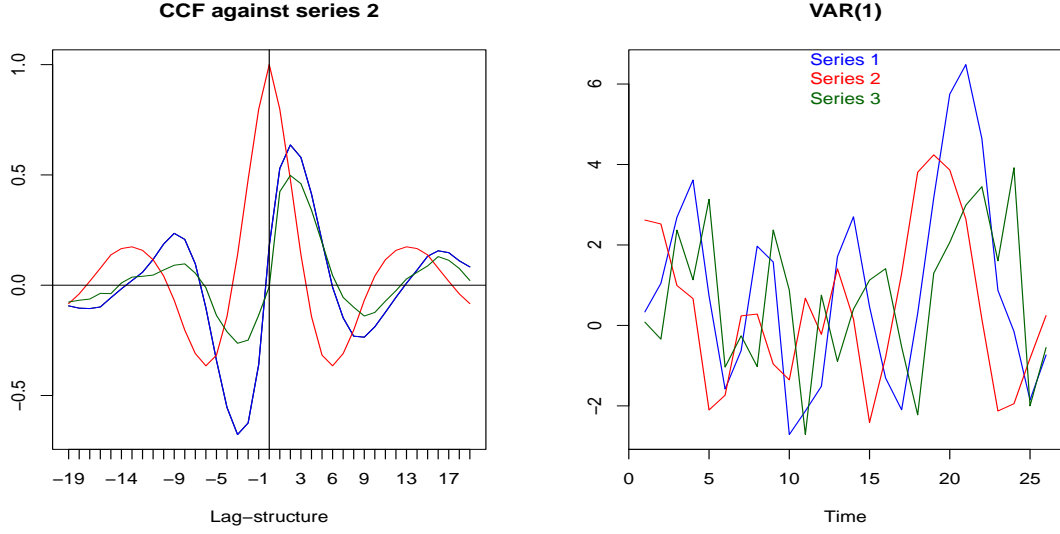
23

Figure 6: Left panel: CCF of second series and first series (blue), ACF of second series (red) and CCF of second series and third series (green): the peaks of the CCFs suggest that the second series x_{2t} is left-shifted or leading. Right panel: a short subsample of x_{it}, $i = 1, 2, 3$: the second series (red) appears left-shifted (leading).

	Sign accuracy	Cor. with data	HT M-SSA	HT of data
1	0.74 (0.74)	0.69 (0.69)	8 (8.01)	3.91 (3.89)
2	0.96 (0.95)	0.99 (0.99)	6 (6)	4.9 (4.88)
3	0.66 (0.66)	0.48 (0.48)	10 (10)	2.12 (2.11)

Table 3: Expected performances of M-SSA (first three columns) and expected HTs of original data $\mathbf{x}_t$ (last column). Sample estimates in parentheses are based on an extensive sample of size $N = 100000$.

Fig.(7) presents a comparison between x_{it} (black) and M-SSA endpoint smoothers y_{it} (cyan), for $i = 1, 2, 3$, thereby confirming the previously discussed findings. Specifically, the smoothing task is considerably less challenging for x_{2t} compared to x_{3t}, as evidenced by the more pronounced discrepancy between the HTs of the original data and the smoother in the latter case. This observation is further supported by the last two columns in Table (3).
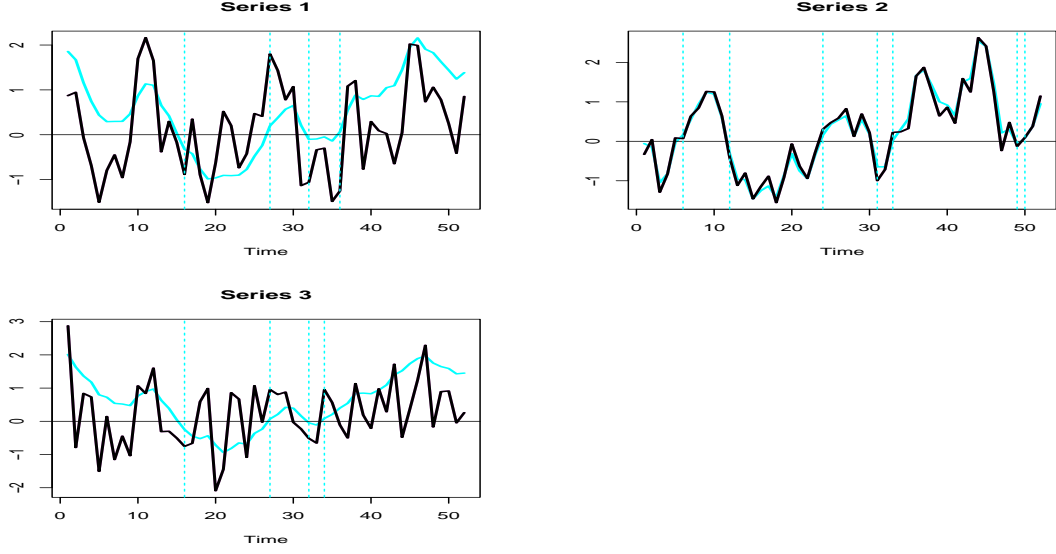
Figure 7: A comparison of original data (black) and SSA-filters (cyan). Zero-crossings of the smoothers are marked by vertical lines.

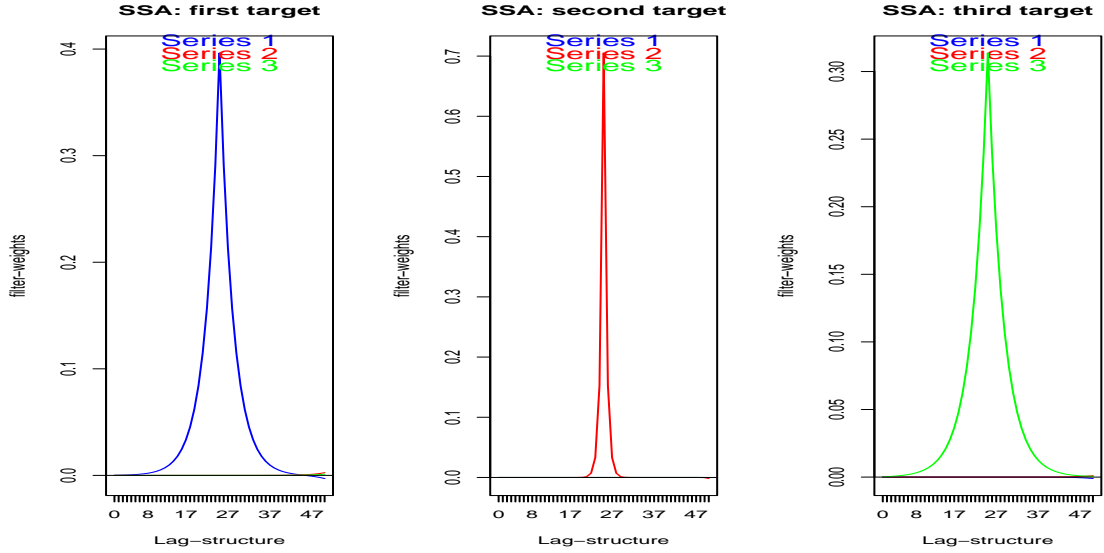To conclude, we also calculate multivariate backcasts for $\delta = -(L-1)/2 = -25$, as illustrated in Fig.(8).

Figure 8: M-SSA backcasts for $\delta = -(L-1)/2$ of the first target series x_{1t} (left), second target x_{2t} (middle) and third target x_{3t} (right). Filter-weights applied to x_{1t} (blue) x_{2t} (red) and x_{3t} (green). The multivariate backcast simplifies to univariate symmetric smoothers for each series.

For each series, the resulting optimal multivariate smoother is virtually identical to a univariate symmetric design. This outcome is intuitively appealing, as the backcast can now utilize data from both sides of the central point, thereby simplifying the smoothing process.

25

4.3 Signal Extraction

The M-SSA criterion is capable of addressing both uni- and multivariate forecasting and smoothing tasks. Here, we consider signal extraction, allowing for both the target and the MA inversion to deviate from the identity, with the latter potentially adopting a multivariate approach, representing the highest level of complexity for M-SSA. To illustrate this, we follow the business-cycle application outlined in Wildi (2024), which applies the HP(14400) filter to the monthly U.S. industrial production index[2]. We extend the scope of this analysis by incorporating an additional leading indicator, the Composite Leading Indicator (CLI) provided by the OECD[3], as depicted in the top left panel of Fig.(9).
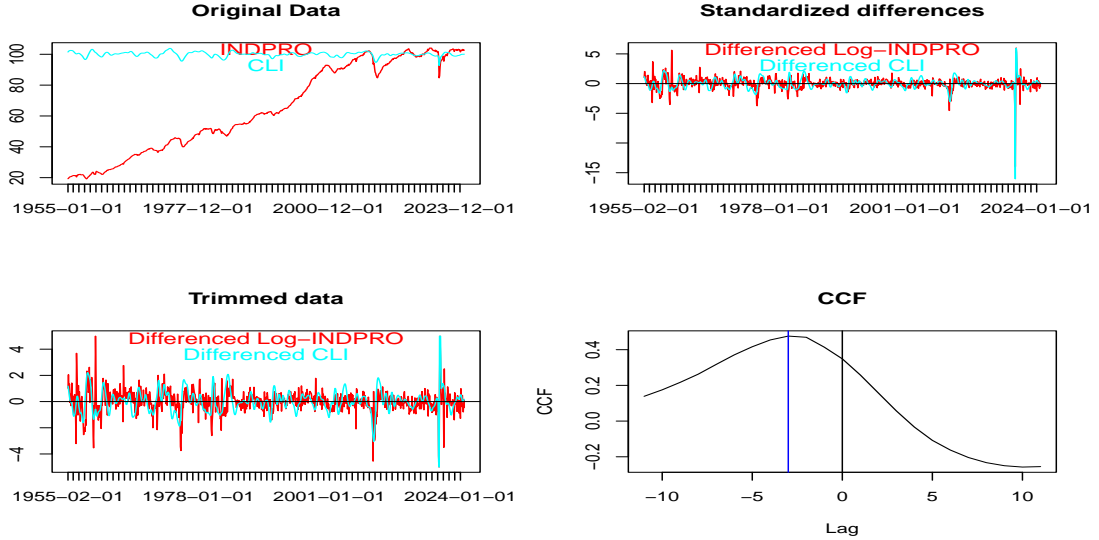

Figure 9: The monthly Industrial Production Index (INDPRO) and Composite Leading Indicator (CLI) are presented in the top left panel, with standardized log-differences displayed in the top right panel. In the bottom left panel, singular pandemic outliers have been removed to ensure that absolute values do not exceed 5σ. The CCF of log-differences is shown in the bottom right panel. The peak of the CCF at a lag of -3 months, suggests an advancement of the CLI by one quarter.

Given that the CLI fluctuates around a fixed level, we now consider the first differences of the log-transformed data. The top right panel of the figure presents the corresponding standardized time series, while the bottom left panel displays the actual data utilized for fitting. In this latter panel, singular pandemic outliers have been removed to ensure that absolute values do not exceed 5σ, a threshold that aligns with extreme observations from earlier historical data. The peak of the CCF in the lower panel indicates that the CLI leads the INDPRO by one quarter. We examine the application of the HP trend filter to the first differences of these series, with a focus on the growth cycle to track U.S. recession episodes as identified by the National Bureau of Economic Research (NBER). In contrast, the conventional output gap is defined as the difference between the actual (original non-stationary) time series and its trend, which addresses deviations from 'potential output'. A comparative analysis of both designs is presented in Wildi (2024).

---

[2]Board of Governors of the Federal Reserve System (US), Industrial Production: Total Index [INDPRO], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/INDPRO, October 23, 2024.

[3]Organization for Economic Co-operation and Development, Leading Indicators OECD: Leading Indicators: Composite Leading Indicator: Amplitude Adjusted for United States [USALOLITOAASTSAM], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/USALOLITOAASTSAM, October 23, 2024.

However, to conserve space, this discussion will utilize the single growth cycle concept to illustrate the M-SSA. We then compare univariate and bivariate designs, highlighting target correlations, holding times (HTs) and leads (left shifts/advancement) or lags (right-shifts/retardation) of the corresponding nowcasts. Specifically, besides the bivariate MSE nowcast we consider the univariate SSA and the classic concurrent HP, as discussed in McElroy (2006), denoted as HP-C, as additional benchmarks. To ensure clarity and maintain focus on the relevant M-SSA topic, we have opted to exclude data revisions and to rely exclusively on the final releases of both indicators. Additionally, the modeling will incorporate all available observations. While alternative shorter subsamples may partially mitigate issues of misspecification, they do so at the risk of overfitting. Nevertheless, the performance of nowcasts remains fairly robust against variations or 'tweaks' in the data sample.

4.3.1 Univariate SSA Benchmark

We begin by deriving a univariate INDPRO-SSA framework, with the filter length set to $L = 201$. This analysis yields the following ARMA(2,1) model for the transformed INDPRO series illustrated in the bottom left panel of Fig.(9):

$$\mathbf{x}_t = 0.01 + 0.96x_{t-1} - 0.16x_{t-2} + \epsilon_t - 0.64\epsilon_{t-1}, \tag{35}$$

where it is noted that the model residuals exhibit white noise characteristics as confirmed by standard diagnostic tests. This model is employed to derive the MA inversion ξ_k for $k = 0, ..., 200$, as illustrated in the left panel of Fig. (10). Our acausal objective $z_{t+\delta}$ utilizes the (truncated) two-sided Hodrick-Prescott filter with $\lambda = 14400$, as illustrated in the middle panel of the figure. The (univariate) MSE nowcast of $z_{t+\delta}$, where $\delta = 0$, is derived following McElroy and Wildi (2016)[4]

$$\hat{\gamma}_{x\delta}(B) = \sum_{k \geq 0} \gamma_{k+\delta} B^k + \sum_{k < 0} \gamma_{k+\delta} \left[\boldsymbol{\xi}(B)\right]_{|k|}^{\infty} B^k \boldsymbol{\xi}^{-1}(B), \tag{36}$$

where B denotes the backshift operator, $\boldsymbol{\xi}(B) = \sum_{k \geq 0} \xi_k B^k$, $\boldsymbol{\xi}^{-1}(B)$ represents the AR-inversion of the data generating process and γ_k, $|k| < \infty$, are the weights of the two-sided HP. The notation $[\cdot]_{|k|}^{\infty}$ signifies the omission of the first $|k| - 1$ lags. Specifically, the derivation of the MSE benchmark follows three distinct steps: first, compute the convolution of the (finite length) two-sided target filter γ_k, $|k| \leq 100$, and the MA inversion ξ_k for $100 \geq k \geq 0$ of the ARMA model; second, truncate this expression by omitting the acausal terms associated with all future innovations ϵ_{t-k} for $k < 0$; lastly, deconvolute the MSE filter from the truncated convolution, as specified in Equation (6) (the M-SSA package provides comprehensive guidance on the sequence of transformations and operations). The resulting filter is depicted in the right panel of Fig. (10) (green line).

The MSE nowcast filter exhibits a lag-one ACF of 0.963, which corresponds to a HT of 11.508. Similarly, the classic HP-C filter has lag-one ACF 0.967, which corresponds to a HT of 12.267. The (univariate) SSA, as illustrated in Fig.(10) (blue line, right panel), is derived from the MSE nowcast, as outlined in Section (2.4). Within the HT constraint, we set $ht_1 = 17.26$, which surpasses the MSE benchmark (and incidentally HP-C) by 50%. As a result, we anticipate that the SSA nowcast will yield a reduction of 33% in zero-crossings over the long term.

---

[4]An open source R-package is provided by McElroy and Livsey (2022) with further discussion available in McElroy, T. (2022).

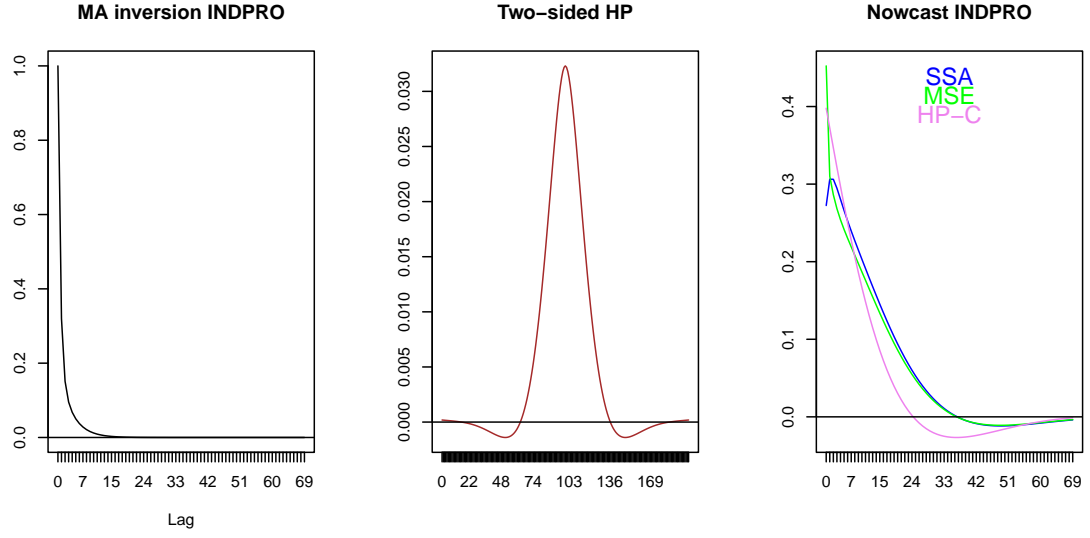| MA inversion INDPRO | Two-sided HP | Nowcast INDPRO |

Figure 10: MA inversion of ARMA-model fitted to INDPRO (left panel). Truncated two-sided HP target filter (middle panel). Univariate MSE (green), HP-C (violet) and SSA nowcast filters (blue) (right panel). Only the first 70 lags are shown for clarity, and all filters are scaled to unit length to facilitate visual comparison.

A comparison of SSA and HP-C nowcasts in Fig.(11) reveals that the SSA exhibits increased smoothness, albeit at the cost of a one-month rightward shift or retardation relative to HP-C. Thus , we anticipate that the bivariate leading indicator design, which incorporates the additional CLI, will mitigate this lag while maintaining comparable smoothness.
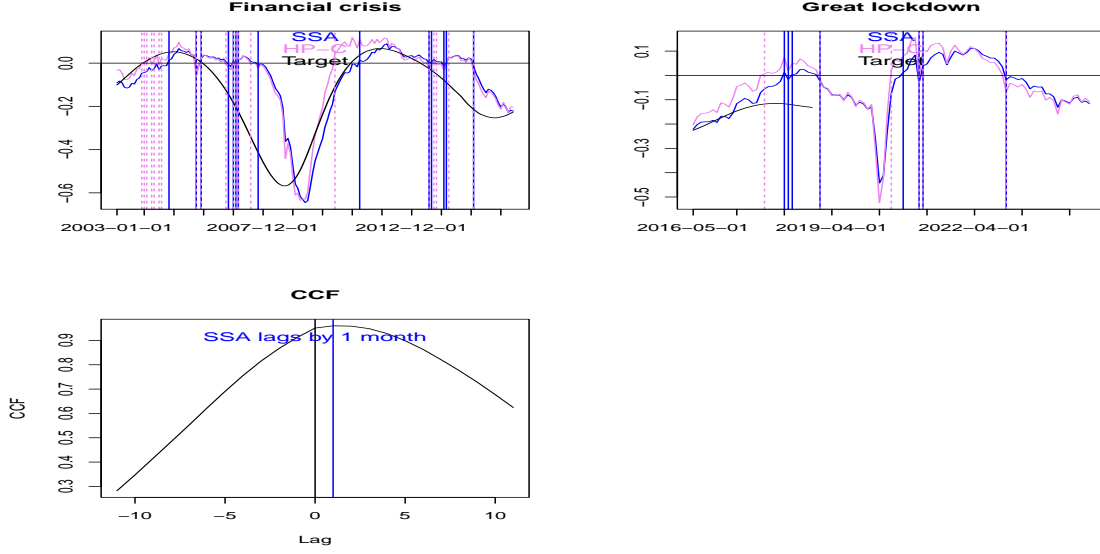
Figure 11: SSA (blue) vs. HP-C (violet), illustrating their performance during the financial crisis (top left) and the pandemic (top right). Zero-crossings are indicated by vertical lines with corresponding colour-coding. The cross-correlation function presented in the bottom panel suggests that SSA exhibits a one-month lag relative to HP-C.

4.3.2 M-SSA: Bivariate Leading Indicator Design

We examine a bivariate design that incorporates the CLI as an additional explanatory variable for the INDPRO. The following VARMA(3,1) model is specified as:

$$\mathbf{x}_t = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0.63 & 0.32 \\ -0.28 & 1.28 \end{pmatrix} \mathbf{x}_{t-1} + \begin{pmatrix} -0.07 & -0.44 \\ -0.05 & -0.36 \end{pmatrix} \mathbf{x}_{t-2} + \begin{pmatrix} 0.02 & 0.3 \\ 0 & 0.09 \end{pmatrix} \mathbf{x}_{t-3} + \boldsymbol{\epsilon}_t + \begin{pmatrix} 0.5 & -0.43 \\ -0.19 & 0.2 \end{pmatrix} \boldsymbol{\epsilon}_{t-1}$$

with the covariance matrix given by $\boldsymbol{\Sigma} = \begin{pmatrix} 0.562 & 0.05414 \\ 0.05414 & 0.1494 \end{pmatrix}$, where the first series corresponds to INDPRO. The bivariate residuals are confirmed to exhibit white noise characteristics according to standard diagnostic tests. Subsequently, we focus attention on a INDPRO nowcast, discarding a corresponding analysis for the CLI. For consistency, we apply the same HT constraint as for the SSA in the previous section. A simulation experiment utilizing an extensive sample of one million

	Cor. HP-C	Cor. M-SSA	Cor. MSE	HT HP-C	HT M-SSA	HT MSE
Expected	0.650	0.736	0.744	11.132	17.263	11.011
Sample	0.649	0.734	0.743	11.120	17.180	10.947

Table 4: Simulation experiment: convergence of sample performances to expected numbers based on a single long sample of length one Million of the bivariate VARMA(3,2)-model.

observations from the aforementioned VARMA process corroborates the convergence of sample performance metrics to their expected values, as detailed in Table (4)It is important to emphasize that the results presented in the table are contingent upon the bivariate VARMA model. Specifically, for the classic HP-C the values would differ if the univariate ARMA model from the previous section were utilized instead. The derivation of these 'true' performance metrics, presented in the first row of the table, is briefly elucidated herein. The expected values are derived from the MA inversions of the filters, as outlined in Section (2.4). Specifically, the target correlation and lag-one ACF of the M-SSA are derived from the expressions in Criterion (7), relying on the MA-inversion

$\mathbf{\Xi}_k$ for $k = 0, \ldots, L-1$ associated with the VARMA process. The corresponding MA weights for the first target, INDPRO, are depicted in the top left panel of Fig. (12). The target correlations and lag-one ACFs for the bivariate MSE and the univariate HP-C benchmarks are obtained using analogous methodologies. In the case of HP-C, the original M-SSA coefficients $\mathbf{B}_k$ are substituted with $\mathbf{HP}_k^{\text{concurrent}} = \begin{pmatrix} hp_k^{\text{concurrent}} & 0 \\ 0 & hp_k^{\text{concurrent}} \end{pmatrix}$, for $k = 0, \ldots, L-1$ in the expressions for the target correlation and the lag-one ACF, where $hp_k^{\text{concurrent}}$ represents the weights of the classic one-sided HP-C. The zero off-diagonal elements of the one-sided $\mathbf{HP}_k^{\text{concurrent}}$, $k \geq 0$, indicate that each target is influenced solely by its own time series. The bivariate MSE predictor can be formulated as described by McElroy and Wildi (2020). Specifically, the expected values presented in the table are contingent upon the MA inversion of the bivariate MSE nowcast, which is conducted in two distinct stages. First, the convolution of $\mathbf{\Xi}_k$ with the (truncated) two-sided HP weights $\mathbf{HP}_k^{\text{two-sided}} = \begin{pmatrix} hp_k^{\text{two-sided}} & 0 \\ 0 & hp_k^{\text{two-sided}} \end{pmatrix}$ for $k = -(L-1), \ldots, L-1$ is computed. Second, all weights associated with future innovations $\boldsymbol{\epsilon}_{t-k}$ for $k < 0$ are excluded from this convolution, thereby retaining only the weights relevant to the causal component of the inverted acausal target[5]. The resultant MA-inverted MSE predictor is then incorporated into the expressions of Criterion (12) (in place of $(\mathbf{b} \cdot \boldsymbol{\xi})_i$) to derive the corresponding lag-one ACF and target correlation. To finish, the HTs reported in Table (4) are derived from Equation (3), utilizing the previously calculated ACFs. It is noteworthy that the effective predictors, such as those illustrated in the top right and bottom panels of Fig. (12), are obtained through the deconvolution of the MA-inverted designs, as described in Equation (6). The M-SSA package provides comprehensive guidance on the sequence of operations and transformations necessary to replicate all results, including those presented in the aforementioned table. It is important to underscore that univariate and bivariate time series models differ substantially in their structure, resulting in variations in the moments or expected performances of the univariate HP-C filter, contingent upon the model employed; the results presented in the table are derived from the VARMA model, while the results obtained from the ARMA model are not included. Nonetheless, the output of the HP-C filter is the same, regardless of the selected model. Additionally, the univariate MSE nowcast and the bivariate MSE nowcast exhibit differences, with the latter demonstrating superior performance. Consequently, we will exclude the univariate MSE from the subsequent benchmark comparisons.

The MA-inversion of the VARMA-model, along with the nowcast weights presented in Fig.(12), indicates that the CLI serves as an important explanatory variable for INDPRO. The rightward shift of the majority of the CLI MA-weights (top left) further substantiates the leading characteristic of the indicator.

---

[5]The MSE nowcast functions as a bivariate design because the VARMA model is employed to project future unobserved data related to the (acausal) left tail of the two-sided HP target. Specifically, for each target series, the deconvolution of the MA-inverted MSE filter yields non-vanishing weights assigned to both time series. In contrast, the deconvolution of the MA-inverted HP-C nowcast reverts to the original univariate HP-C filter.
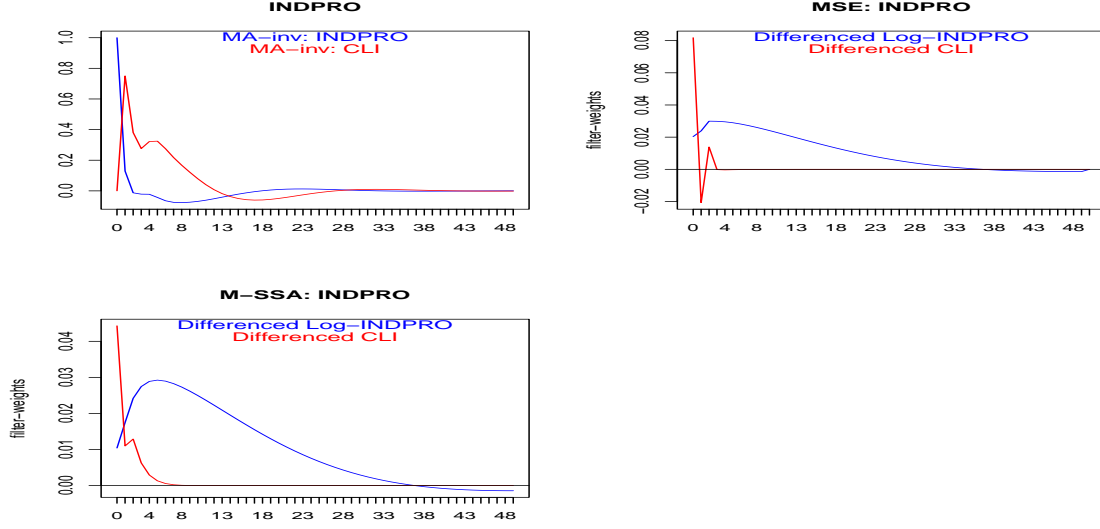
Figure 12: MA-inversion of VARMA(3,1) model (top left). MSE (top right) and M-SSA nowcasts (bottom) for INDPRO: all series are truncated to length 50 to facilitate visual comparison.

The sample and expected performances of the classic univariate HP-C filter and the univariate SSA, alongside the bivariate MSE and M-SSA nowcasts are compared in Tables (5) (which includes target correlations pertaining to the two-sided HP) and (6) (which presents HTs), with standard errors indicated in parentheses. All expected values are based on the bivariate VARMA model, except for the SSA values (in the second columns of the tables), which are derived from the univariate ARMA model. Consequently, direct comparisons between these two should be approached with caution. The sample estimates are derived from data spanning from 1955-02-01 to 2024-09-01, resulting in a net number of observations of $N = 605$[6]. A comparison of expected and sample HTs indicates that the latter are consistently larger, with a statistically significant finding for the MSE nowcast, which may suggest potential model misspecification[7]. The SSA and M-SSA designs exhibit similar HTs and greater smoothness compared to the MSE, as evidenced by both expected and sample performances, further corroborated by the corresponding filter outputs in Fig. (13). The M-SSA demonstrates superior performance compared to the traditional univariate HP-C filter in terms of target correlation and HT. Additionally, the CCF depicted in Fig. (14) (bottom right panel) shows that M-SSA and HP-C are coincident, unlike the SSA discussed in the previous section, which displayed a lagging behavior. The figure further highlights a two-month lead of the M-SSA design over the univariate SSA (top right). The MSE nowcast reflects enhanced expected an sample target correlations, noting that the expected value of the (univariate) SSA cannot be directly compared due to differing underlying models.

---

[6]We utilize the full available common sample of INDPRO and CLI to mitigate the impact of excessive variance in the sample HT. The net sample length is calculated as $T - L - L^{left-tail} = 836 - 201 - 30 = 605$, where L represents the filter length (initialization for all designs) and $L^{left-tail}$ denotes the length of the left 'acausal' side of the target, thereby restricting comparisons to time points $t \leq T - L^{left-tail}$ near the end of the sample.

[7]Alternative models employing shorter subsamples may partially alleviate this issue without impacting our primary conclusions; thus, they are not included in this discussion.

	Cor. HP-C	Cor. SSA	Cor. M-SSA	Cor. MSE
Expected	0.65	0.755	0.736	0.744
Sample	0.735(0.074)	0.717(0.085)	0.77(0.077)	0.791(0.073)

Table 5: Expected and sample target correlations of univariate HP-C and SSA (first two columns) and bivariate M-SSA and MSE nowcasts, with standard deviations in parentheses.

	HT HP-C	HT SSA	HT M-SSA	HT MSE
Expected	11.132	17.263	17.263	11.011
Sample	15.512(2.183)	18.508(6.629)	24.462(4.922)	19.875(3.081)

Table 6: Expected and sample HTs of univariate HP-C and SSA (first two columns) and bivariate M-SSA and MSE nowcasts, with standard deviations in parentheses.
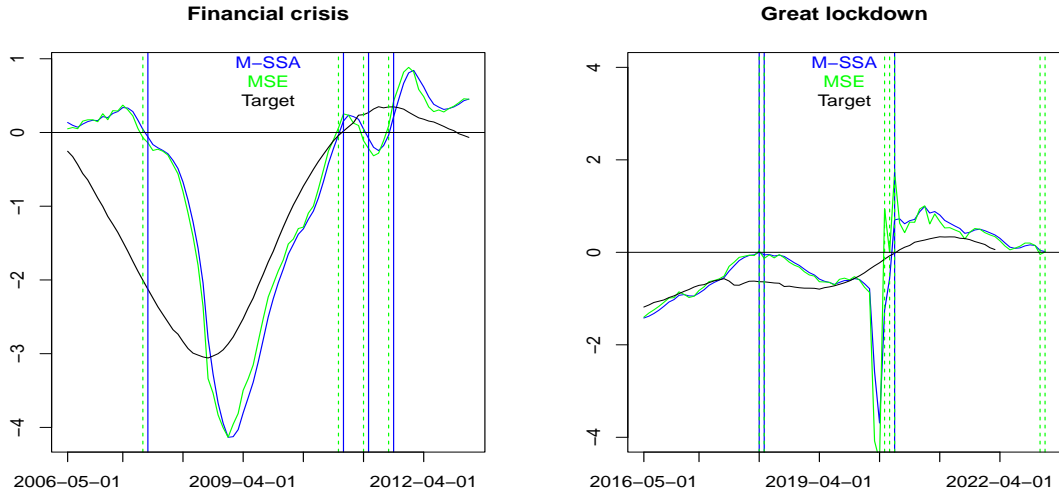

Figure 13: M-SSA (blue), MSE (green) and acausal target (black) with zero-crossings of M-SSA (vertical blue lines) and MSE (vertical green lines).
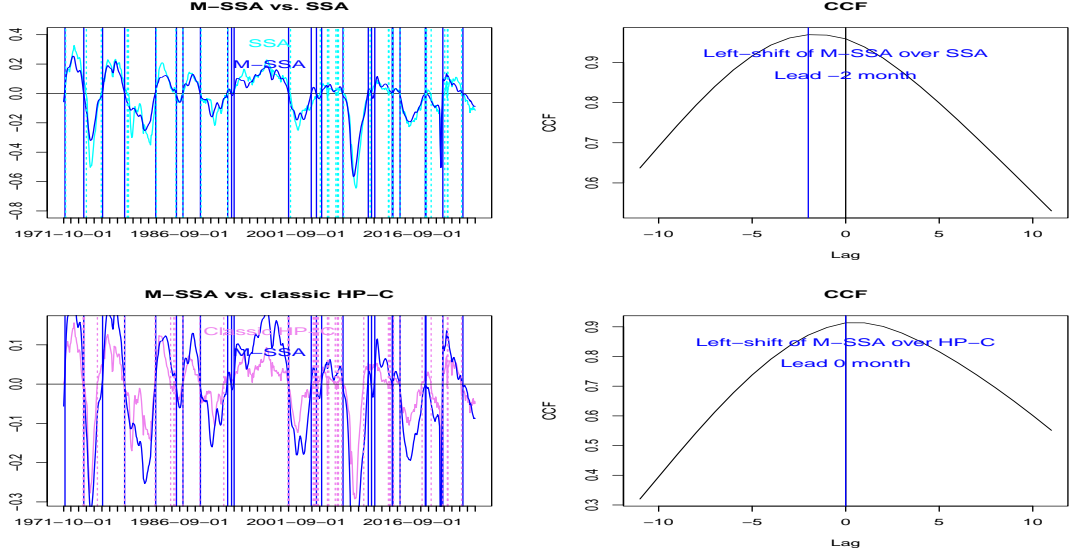
Figure 14: INDPRO nowcasts: indicators (left) and CCF (right); M-SSA (blue) vs. SSA (cyan) (top left) and M-SSA (blue) vs. classic HP concurrent (violet) (bottom left). Vertical lines indicating zero-crossings are color-coded accordingly. The CCFs on the right demonstrate that M-SSA leads both the SSA and the classic HP concurrent filter by two months.

To conclude, it is interesting to examine the filtered components of the INDPRO M-SSA nowcast, which are derived from the sub-filters applied to both INDPRO and CLI. Figure (15) illustrates these components alongside the aggregate for both the financial crisis and the 'great lockdown'. Prior to and following the crises, the dynamics of the CLI component are relatively weak, and the nowcast (in blue) aligns with the INDPRO component. However, during the crisis, the CLI exhibits increased dynamics, leading to a divergence of the aggregate M-SSA from the INDPRO component. Notably, the lead introduced by the CLI results in a leftward shift of the aggregate nowcast in relation to the INDPRO component.
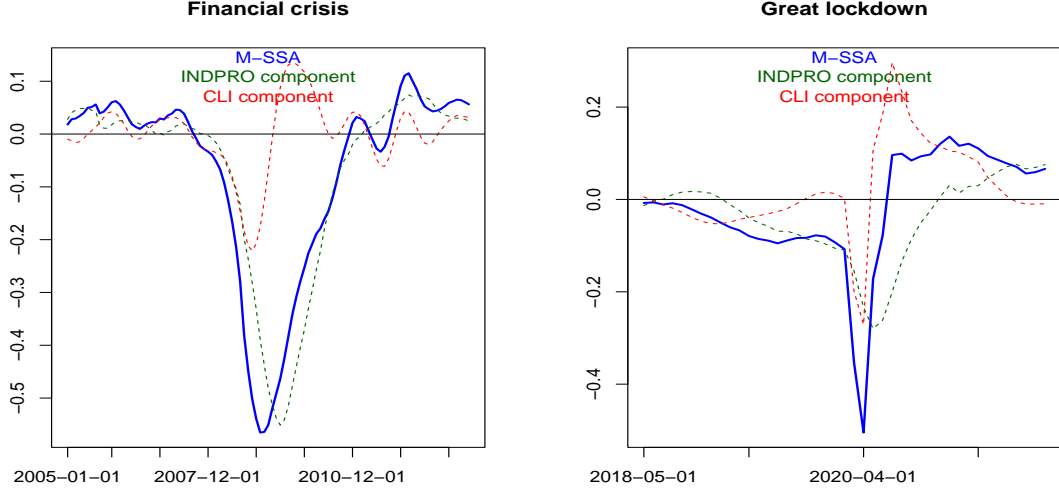
Figure 15: The INDPRO M-SSA nowcast is represented by the blue solid line, along with its components derived from the sub-filters applied to INDPRO (green dashed line) and the filtered CLI (red dashed line): financial crisis (left) and 'great lockdown' (right).

5   Conclusion

We propose a multivariate extension of the SSA, termed M-SSA, as introduced in Wildi (2024). M-SSA represents an innovative approach to prediction and smoothing, highlighting target correlation and the sign accuracy of the predictor while adhering to a novel HT constraint. The conventional MSE approach corresponds to unconstrained M-SSA optimization, subject to a specific scaling factor. However, our proposed criterion captures more nuanced performance metrics related to accuracy and smoothness characteristics. In its primal formulation, M-SSA seeks to optimally track the target while enforcing noise suppression; conversely, in its dual formulation, the predictor minimizes zero-crossings for a predetermined level of tracking accuracy.

The M-SSA predictor is both interpretable and attractive due to its inherent simplicity, as it integrates fundamental concepts of prediction, such as sign accuracy, mean squared error, and smoothing requirements. The predictor operates under a simple time-reversible unstable difference equation. Importantly, the stability of the predictor is maintained through the presence of implicit 'zero' boundary conditions.

The M-SSA framework can effectively address forecasting, smoothing, and signal extraction applications, depending on whether its target specification is causal or non-causal, as well as whether it is all-pass or not. This methodology enables the customization of benchmark predictors in terms of smoothness and accuracy performance. Our applications in forecasting, smoothing, and signal extraction demonstrate the customization of MSE and HP benchmarks. Furthermore, our M-SSA package expands this analysis to incorporate additional contemporary business-cycle tools. Notably, the bivariate signal extraction application highlights the significance of leading indicators within a nowcasting framework.

References

[1] Cerreia-Vioglio, S., Ortu, F., Severino, F. et al (2023) Multivariate Wold decompositions: a Hilbert A-module approach.*Decisions Econ Finan* **46**, 45–96. https://doi.org/10.1007/s10203-023-00392-3

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

[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] Horn, R., Johnson, C. (1991) Topics in Matrix Analysis. Cambridge: Cambridge University Press.

[7] Kamber, G. Morley, J. Wong, B. (2024). Trend-Cycle Decomposition in the Presence of Large Shocks. *CAMA Working Papers 2024-24.* Centre for Applied Macroeconomic Analysis, Crawford School of Public Policy, The Australian National University, revised Aug 2024.

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

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

[10] McElroy, T. (2006) Exact Formulas for the Hodrick-Prescott Filter. *Research report series (Statistics 2006-9). U.S. Census Bureau* .

[11] McElroy, T. (2022) Casting vector time series: algorithms for forecasting, imputation, and signal extraction. *Electron. J. Statist.* **16 (2)**, 5534-5569. DOI: 10.1214/22-EJS2068

[12] McElroy, T.S. and Livsey, J.A. (2022) Ecce Signum: An R Package for Multivariate Signal Extraction and Time Series Analysis *arXiv:2201.02148*, https://arxiv.org/abs/2201.02148

[13] McElroy, T. and Wildi , M. (2019) The trilemma between accuracy, timeliness and smoothness in real-time signal extraction. *International Journal of Forecasting* **35 (3)**, 1072-1084.

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

[15] Ravn M. Uhlig H. (2002) On Adjusting the Hodrick-Prescott Filter for the Frequency of Observations. *The Review of Economics and Statistics* **84** (2), 371-376.

[16] Sophocles, J. Orfanidis (2007) Optimum signal processing. McGraw-Hill.

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

[18] Wildi, M. (2024) Business Cycle Analysis and Zero-Crossings of Time Series: a Generalized Forecast Approach. *J. Bus. Cycle Res.* https://doi.org/10.1007/s41549-024-00097-5.

[19] Wildi, M. (2025) Sign Accuracy, Mean-Square Error and the Rate of Zero Crossings: a Generalized Forecast Approach. *Submitted for publication.*