Prediction, Smoothing and a Trilemma

Marc Wildi

February 14, 2024

Abstract

We propose a generic forecast approach, called Simple Sign Accuracy (SSA), which merges various facets of the prediction problem in terms of mean-square error, sign accuracy, zero crossings and smoothness. The latter is obtained in terms of a novel holding-time constraint which conditions the frequency of sign-changes by the predictor or the forecast error. We obtain a solution to the optimization problem and we derive the distribution of the predictor under fairly general conditions, including a specific treatment of singular cases. In addition to being a self-contained and original forecast approach, we show that SSA can be plugged on an existing benchmark, in view of modifying its characteristics effectively. Moreover, the resulting hybrid predictor inherits interpretations, such as an economic meaning, directly from the benchmark, due to optimal transfer. In its simplest expression, our approach can be linked to smoothing and we compare the new zero-crossing constraint to classic curvature penalty. Finally, we show that SSA sits on top of a generic prediction trilemma that addresses important user-priorities, relevant to nowcasting and forecasting applications, and we propose to customize predictors or smoothers for specific uses.

1 Introduction

Time series applications such as, e.g., forecasting or (real-time) signal extraction formalize attempts to educe information that is not yet readily available or accessible, from present and past data. Typically, optimality concepts rely on prediction accuracy or, more precisely, on the minimization of a distance measure between a target, for example a future observation or a (future or current) trend component, and the predictor. While this proceeding seems uncontroversial, in principle, we argue that alternative characteristics of a predictor might draw attention such as the smoothing capability, i.e., the extent by which undesirable 'noisy' components can be suppressed, or timeliness, as measured by lead (advancement,left-shift of a time series) or lag (retardation, right-shift) properties, or sign accuracy and zero-crossings, as measured by the ability to predict the correct sign of the target. We here propose a generic forecast approach which, under suitable assumptions, merges sign accuracy and mean-square error (MSE) performances subject to a holding time constraint which determines the expected number of zero-crossings of the predictor (or of the forecast error) in a fixed time interval. McElroy Wildi (2019) propose an alternative methodological framework for addressing specific facets of the forecast problem but their approach does not accommodate for zero-crossings explicitly which may be viewed as a shortcoming in some applications.

The analysis of zero-crossings of a time series has been pioneered by Rice (1944) who derives a link between the autocorrelation function (ACF) of a zero-mean stationary Gaussian process and its expected number of crossings in a fixed interval. Interestingly, sign changes of successively differenced processes can be informative about the entire autocorrelation sequence and thus the spectrum of a stationary time series, see Kedem (1986). A theoretical overview is provided by Kratz (2006). Applications have been proposed in the field of exploratory and inferential statistics, see Kedem (1986) and Barnett (1996) and are numerous in electronics and image processing, process discrimination, or pattern detection in speech, music, or radar screening. However, while most applications concern the analysis of current or past events, we here emphasize mainly a prospective prediction perspective.

In addition to being a self-contained and original forecast approach, SSA can be plugged on an existing benchmark predictor to modify some of its characteristics in terms of timeliness, smoothness or MSE performances. Specifically, we show that these three terms constitute a forecast trilemma on top of which SSA can trigger priorities by means of hyperparameters. Wildi (2024) proposes an application of SSA to a (real-time) business-cycle analysis, but the chosen treatment remains largely informal. We here fill this gap by providing a complete formal treatment, including regular, singular and boundary cases, a discussion of numerical aspects as well as a derivation of the sample distribution of the predictor together with a comprehensive illustration of technical features and peculiarities of the approach. All empirical examples can be replicated in an open source SSA-package[1]. The proposed forecast approach is generic and can be extended to alternative signal specifications, predictors or data, for any forecast or backcast horizons: for illustration, in addition to the HP filter, the SSA-package proposes applications to Hamilton's regression filter, Hamilton (2018), and to the Baxter-King (BK) bandpass filter, Baxter and King (1999).

2 Simple Sign-Accuracy (SSA-) Criterion

2.1 Introduction

Consider $z_t = \sum_{k=-\infty}^{\infty} \gamma_k \epsilon_{t-k}$ where $\epsilon_j, j \in \mathbb{Z}$, is standardized white noise and $\boldsymbol{\gamma} := (\gamma_k), k \in \mathbb{Z}$, is a (real) square summable sequence whose weights γ_k are applied to x_{t-k} so that z_t is a stationary zero-mean process with variance $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We consider estimation of $z_{t+\delta}$, $\delta \in \mathbb{Z}$, referred to as the *target*, based on the predictor $y_t := \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$, where b_k are the coefficients of a one-sided causal filter of length L. This problem is commonly referred to as fore-, now- or backcast, depending on $\delta > 0$, $\delta = 0$ or $\delta < 0$, respectively. Consider the following optimization problem

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

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

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

is of dimension $L \cdot L$ and l is an arbitrary scaling. Criterion (1) is referred to as *simple sign accuracy* or SSA criterion, its solution is denoted by SSA(ρ_1, δ); the constraints $\mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1$ and $\mathbf{b}'\mathbf{b} = l$ are referred to as holding time (ht) and length constraints, respectively, see Wildi (2024). To simplify terminology, we now merge the concepts of filter outputs and filter weights such that, e.g., $y_{t,MSE} := \boldsymbol{\gamma}_\delta' \boldsymbol{\epsilon}_t$ or $\boldsymbol{\gamma}_\delta$ will both be referred to as MSE predictor and similarly $y_t = \mathbf{b}\boldsymbol{\epsilon}_t$ or $\mathbf{b}$ are the SSA predictor, where $\boldsymbol{\epsilon}_t := (\epsilon_t, ..., \epsilon_{t-(L-1)})$. Note that the MSE-predictor $\boldsymbol{\gamma}_\delta$ stands for the proper target $\boldsymbol{\gamma}$ in Criterion (1) and we now assume $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$: $\mathbf{b}$ (or $y_t := \mathbf{b}\boldsymbol{\epsilon}_t$) can be interpreted as a (constrained) predictor for $z_{t+\delta}$; alternatively, $\mathbf{b}$ can be viewed as a 'smoother' for $\boldsymbol{\gamma}_\delta$, see Section (6); in this sense, the SSA criterion merges prediction and smoothing and the particular objective function can retain pertinence outside of a classic prediction or MSE paradigm. Also,

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

$\mathbf{b}'\mathbf{M}\mathbf{b}/l = \mathbf{b}'\mathbf{M}\mathbf{b}/\mathbf{b}'\mathbf{b} =: \rho(y, y, 1)$ is the lag-one autocorrelation (ACF) of y_t and the objective function $\mathbf{b}'\boldsymbol{\gamma}_\delta$ is proportional to $\rho(y, z, \delta) := \mathbf{b}'\boldsymbol{\gamma}_\delta/\sqrt{l\boldsymbol{\gamma}'\boldsymbol{\gamma}}$, the target correlation of y_t with $z_{t+\delta}$, or to $\mathbf{b}'\boldsymbol{\gamma}_\delta/\sqrt{l\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, the correlation of y_t with $y_{t,MSE}$: maximizing either of these objective functions maximizes the other ones, too and therefore Criterion (1) is equivalent to

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

Finally, an extension to $\tilde{z}_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k}$ where x_t is an autocorrelated stationary or non stationary integrated process is proposed in Section (5).

3 Solution to the SSA-Criterion

The eigenvectors $\mathbf{v}_j$ of $\mathbf{M}$ are the Fourier vectors $\mathbf{v}_j = \left(\sin(k\omega_j)/\sqrt{\sum_{k=1}^{L} sin(k\omega_j)^2}\right)_{k=1,...,L}$ with adjoined eigenvalues $\lambda_j = \cos(\omega_j)$ computed at the discrete Fourier frequencies $\omega_j = j\pi/(L + 1)$, $j = 1, ..., L$, see Anderson (1975): we normalize the eigenvectors to constitute an orthonormal basis of $\mathbb{R}^L$, which will simplify subsequent notation.

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

Proof: Let $\mathbf{b}'\mathbf{b} = 1$ and $\rho(y, y, 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \mathbf{b}'\mathbf{M}\mathbf{b}$. A stationary point of $\rho(y, y, 1)$ is found by equating the derivative of the Lagrangian $\mathfrak{L} = \mathbf{b}'\mathbf{M}\mathbf{b} - \lambda(\mathbf{b}'\mathbf{b} - 1)$ to zero i.e. $(\mathbf{M} + \mathbf{M}')\mathbf{b} = 2\mathbf{M}\mathbf{b} = 2\lambda\mathbf{b}$. We deduce that $\mathbf{b}$ is a stationary point if and only if it is an eigenvector of $\mathbf{M}$. Then

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

for some $i \in \{1, ..., L\}$ and therefore $\rho(y, y, 1)$ must be the corresponding eigenvalue, as claimed. Since the unit-sphere is free of boundary-points, we conclude that the extremal values $\rho_{min}(L)$, $\rho_{max}(L)$ must be stationary points so that $\rho_{min}(L) = -\cos(\pi/(L+1))$ and $\rho_{max}(L) = \cos(\pi/(L+1))$ and the boundaries are obtained for $\mathbf{b} := \mathbf{v}_1$ and $\mathbf{b} := \mathbf{v}_L$, respectively. $\square$

Consider the spectral decomposition of the target $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$

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

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

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

A proof follows directly from Proposition (1), noting that $\mathbf{b}_1'\boldsymbol{\gamma}_\delta = \text{sign}(w_1)w_1 > 0$ and $\mathbf{b}_L'\boldsymbol{\gamma}_\delta = \text{sign}(w_L)w_L > 0$ (maximization).

Theorem 1. *Consider the SSA Criterion* (1) *under the posited assumptions about* $z_{t+\delta}$ *and assume the following set of regularity assumptions hold:*

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

Then:

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

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

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

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

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

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

3. *The derivative* $d\rho(\nu)/d\nu$ *is strictly negative, for* $\nu \in \{x||x| > 2\rho_{max}(L)\}$*, and* $max_{\nu<-2\rho_{max}(L)}\rho(\nu) = min_{\nu>2\rho_{max}(L)}\rho(\nu) = \rho_{MSE}$*, where* ρ_{MSE} *denotes the lag-one ACF of* $\boldsymbol{\gamma}_\delta$.

4. *For* $\nu \in \{x||x| > 2\rho_{max}(L)\}$

$$-sign(\nu)\frac{d\rho(y(\nu), z, \delta)}{d\nu} = \frac{1}{(\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1}{}'\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta)^{3/2}\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}}\frac{d\rho(\nu)}{d\nu} < 0 \tag{6}$$

A complete proof of the theorem is provided in the appendix. The case of incomplete spectral support is addressed in the following corollary.

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

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

$$\mathbf{b}(\nu) = D \sum_{i \in NZ} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \tag{7}$$

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

$$\rho(\nu) = \frac{\sum_{i \in NZ} \frac{\lambda_i w_i^2}{(2\lambda_i - \nu)^2}}{\sum_{i \in NZ} \frac{w_i^2}{(2\lambda_i - \nu)^2}} =: \frac{M_1}{M_2} \tag{8}$$

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

2. *Let $\nu = \nu_{i_0} := 2\lambda_{i_0}$ where $i_0 \notin NZ$ with adjoined rank-defficient $\boldsymbol{\nu}_{i_0} = 2\mathbf{M} - \nu_{i_0}\mathbf{I}$. Consider $\mathbf{b}(\nu_{i_0})$, $\rho(\nu_{i_0})$ and M_{i_01}, M_{i_02} as defined in the previous assertion. In this case, $\mathbf{b}(\nu_{i_0})$ can be 'spectrally completed' as in*

$$\mathbf{b}_{i_0}(\tilde{N}_{i_0}) := \mathbf{b}(\nu_{i_0}) + D\tilde{N}_{i_0}\mathbf{v}_{i_0} \tag{9}$$

with lag-one acf

$$\rho_{i_0}(\tilde{N}_{i_0}) = \frac{M_{i_01} + \lambda_{i_0}\tilde{N}_{i_0}^2}{M_{i_02} + \tilde{N}_{i_0}^2} \tag{10}$$

If i_0 is such that $0 < \rho(\nu_{i_0}) = \frac{M_{i_01}}{M_{i_02}} < \rho_1 < \lambda_{i_0}$ or $0 > \rho(\nu_{i_0}) = \frac{M_{i_01}}{M_{i_02}} > \rho_1 > \lambda_{i_0}$, then

$$\tilde{N}_{i_0} = \pm\sqrt{\frac{\rho_1 M_{i_02} - M_{i_01}}{\lambda_{i_0} - \rho_1}} \tag{11}$$

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

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

A proof of the corollary is provided in the appendix.

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

A proof follows directly from Theorem (1), noting that the scaling $s = \sqrt{l/\mathbf{b}'\mathbf{b}}$ interferes neither with the objective function nor with the holding time constraint. In particular, Assertion (3) warrants uniqueness for $\nu \in \{x||x| > 2\rho_{max}(L)\}$.

Assertion (3) ensures swift numerical optimization with the approximation error being of order $1/2^n$, where n is the number of iterations. However, to enlighten the structure of the SSA problem further we provide an exact closed-form solution in the case of an AR(1) target $\gamma_{k+\delta} = Ca_1^k$.

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

1. *The MSE-estimate $\boldsymbol{\gamma}_\delta$ corresponds to a stationary AR(1) i.e. $\gamma_k \propto \lambda^k$, $k = 0, ..., L-1$ with stable root $\lambda \neq 0$ (exponential decay)*
2. *$|\nu| > 2$ (typical 'non unit-root' case). In this case $\nu = \lambda_{1\rho_1} + 1/\lambda_{1\rho_1}$ where $\lambda_{1\rho_1} \in \mathbb{R} \setminus \{0\}$.*
3. *$\lambda \neq \lambda_{1\rho_1}$ (non-singular case)*
4. *$\lambda_{1\rho_1}, \lambda$ and L are such that $\max(|\lambda_{1\rho_1}|^{2k}, |\lambda|^{2k})$ is negligible for $k > L$ (sufficiently fast decaying filter-weights).*

Then the optimal invertible root $\lambda_{1\rho_1}$ in $\nu = \lambda_{1\rho_1} + 1/\lambda_{1\rho_1}$ is given by (the real-valued)

$$\lambda_{1\rho_1} = -\frac{1}{3c_3}\left(c_2 + C + \frac{\Delta_0}{C}\right) \tag{12}$$

where

$$\begin{aligned} C &= \sqrt[3]{\frac{\Delta_1+\text{sign}(\Delta_1)\sqrt{\Delta_1^2-4\Delta_0^3}}{2}} \\ \Delta_0 &= c_2^2-3c_3c_1 \\ \Delta_1 &= 2c_2^3-9c_3c_2c_1+27c_3^2c_0 \end{aligned} \tag{13}$$

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

$$\begin{aligned} c_3 &= \lambda^{2\delta-2}-\lambda^{2\delta+2}+\lambda^{2+\delta}-\rho_1\lambda^{2\delta-1} \\ c_2 &= -(\lambda^{1+\delta}+\lambda^{2\delta-1}(1-\lambda^2))-\rho_1\Big(-2\lambda^{2\delta}+\lambda^{2\delta-2}\Big) \\ c_1 &= -(\lambda^{2+\delta}+\lambda^{2\delta}(1-\lambda^2))-\rho_1\Big(\lambda^{2\delta+1}-2\lambda^{2\delta-1}\Big) \\ c_0 &= \lambda^{1+\delta}-\rho_1\lambda^{2\delta} \end{aligned}$$

The SZC-estimate $\mathbf{b}$ *is then uniquely determined in closed-form by 4, down to the correct sign which leads to a positive criterion value* $\mathbf{b}'\boldsymbol{\gamma}_\delta \geq 0$.

A proof of the corollary is postponed to the appendix. The distribution of the SSA-predictor is addressed in the following corollary.

Corollary 5. *Let all regularity assumptions of theorem 1 hold and let $\hat{\boldsymbol{\gamma}}_\delta$ be a finite-sample estimate of the MSE-predictor $\boldsymbol{\gamma}_\delta$ with mean $\boldsymbol{\mu}_{\gamma_\delta}$ and variance $\boldsymbol{\Sigma}_{\gamma_\delta}$. Then mean and variance of the SSA-predictor $\hat{\mathbf{b}}$ are*

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

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

The proof follows directly from Equation 4 and we refer to standard textbooks for a derivation of mean, variance and (asymptotic) distribution of the MSE-estimate under various assumptions about x_t, see Brockwell and Davis (1993). The next corollary derives a dual interpretation of the SSA-predictor.

Corollary 6. *Let all regularity assumptions of Theorem (1) hold and let $y_t(\nu_1)$ denote the SSA-solution for some $\nu_1 > 2\rho_{max}(L)$. Set $\rho_{\nu_1,\delta} := \rho(y(\nu_1), z, \delta) > 0$ and consider the dual optimization problem*

$$\left.\begin{array}{c} \max_{\mathbf{b}} \rho(y,y,1) \\ \rho(y,z,\delta) = \rho_{\nu_1,\delta} \end{array}\right\} \tag{14}$$

If the search for ν can be restricted to the set $\{\nu||\nu| > 2\rho_{max}(L)\}$ then $y_t(\nu_1)$ is also the solution to the dual problem. If $\nu_1 < -2\rho_{max}(L)$, then $y_t(\nu_1)$ is the solution to the dual problem if minimization is substituted for maximization in the criterion.

Proof: Consider first the case $\nu_1 > 2\rho_{max}(l)$. The Lagrangian Equation (38) does not discern constraint and objective: after suitable re-scaling of multipliers, the problem specified by Criterion (14) leads to the same functional form $\mathbf{b} = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta$ of its solution[2]. The only difference is that ν in Criterion (14) must be selected such that $\rho(y(\nu), z, \delta) = \rho_{\nu_1,\delta}$. If the search can be restricted to $\nu \in \{x|x > 2\rho_{max}(L)\}$, then by Assertion (4) the solution to the primal problem

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

is also a solution to the dual problem, due to strict monotonicity of $\rho(y(\nu), z, \delta)$. The extension to $\nu \in \{x||x| > 2\rho_{max}(L)\}$ follows from Assertion (3) which affirms that $\rho(\nu) < \rho(\nu_1)$ if $\nu < -2\rho_{max}(L)$. A similar reasoning applies if $\nu_1 < -2\rho_{max}(L)$, noting that maximization must be replaced by minimization in the dual Criterion (14) (because $\rho(\nu) > \rho(\nu_1)$ if $\nu > 2\rho_{max}(L)$).$\square$

We now interpret the obtained SSA solution and we shall see that the restriction $|\nu| > 2\rho_{max}(L)$ in the previous Corollaries (3) and (6) is not a limitation since in applications, typically the more stringent condition $|\nu| > 2$ (in fact, $\nu > 2$) applies.

4 Interpretation

4.1 A Frequency-Domain Analysis

We now briefly interpret the SSA-solution in the frequency domain. Formally, the SSA AR(2) filter (SSA-AR(2) for short) in the difference equation (40) has transfer function $\Gamma_{AR(2)}(\nu, \omega) = \frac{1}{\exp(-i\omega)-\nu+\exp(i\omega)} = \frac{1}{2\cos(\omega)-\nu}$. Let then $\mathbf{\Gamma}_{AR(2)}(\nu)$ denote the vector of transfer function ordinates of SSA-AR(2) evaluated at the Fourier frequencies $\omega_j = j\pi/(L+1)$, $j = 1, ..., L$. Equation (42) then implies

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

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

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

where $\mathbf{e}_i$ is the i-th unit vector. The latter expression can be interpreted as discrete Fourier transform (SSA-DFT) and its square as SSA-periodogram of the predictor. Also

$$\mathbf{b}'(\nu)\mathbf{b}(\nu) = E\left[\mathbf{b}(\nu)'\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\mathbf{b}(\nu)\right] = E\left[\mathbf{b}(\nu)'\mathbf{V}\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'\mathbf{V}'\mathbf{b}(\nu)\right] = D(\nu, l)^2\sum_{i=1}^{L}\left(\frac{w_i}{2\lambda_i - \nu}\right)^2$$

is Parseval's identity and $\left(\frac{w_i}{2\lambda_i-\nu}\right)^2$ measures the contribution of $\mathbf{v}_i$ to the variance of the predictor up to the (arbitrary) scaling $D(\nu, l)^2$. For $\nu \leq -2$, SSA-AR(2) is a highpass with a peak of its transfer (or amplitude) function at frequency π; if $\nu \to -\infty$ then $D(\nu, l)\mathbf{\Gamma}_{AR(2)}(\nu) \to \mathbf{1}/\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, where $\mathbf{1}$ is a vector of length L of ones, and $\mathbf{b}(\nu) \to \boldsymbol{\gamma}_\delta/\sqrt{\boldsymbol{\gamma}_\delta'\boldsymbol{\gamma}_\delta}$, the normalized MSE predictor. The highpass favors noise leakage as requested when $ht_1 < ht_{MSE}$ in the holding time constraint. For $\nu \geq 2$, SSA-AR(2) is a lowpass with a peak of its transfer or amplitude function at frequency zero: the filter damps high-frequency noise as requested when $ht_1 > ht_{MSE}$ in the holding time constraint. For $-2 < \nu < 2$, SSA-AR(2) is a bandpass design with a peak of its transfer or amplitude function at frequencies $\omega = \pm\arccos(\nu/2)$.

For illustration, we apply the SSA criterion to the quarterly HP filter with parameter $\lambda = 1600$, see Hodrick and Prescott (1997), with two-sided (bi-infinite symmetric) target γ_k displayed in Figure (1). We aim at approximating the HP target $z_{t+\delta}$ for $\delta = 0$ (nowcasting) by a nowcast y_t based on a one-sided filter b_k, $k = 0, ..., 100$, of length $L = 101$. The MSE nowcast $\boldsymbol{\gamma}_0$ corresponds to the right tail of the two-sided filter and has lag-one ACF $\rho_{MSE} = 0.926$. We compute two SSA nowcasts, imposing lag-one ACFs of $0.97 > \rho_{MSE}$ (smoothing) and $0.8 < \rho_{MSE}$ (un-smoothing) with resulting $\nu_1 = 2.44 > 2$ and $\nu_2 = -2.42 < -2$, see Fig.(1). Optimal smoothing and un-smoothing are obtained by lowpass ($\nu_1 > 2$) and highpass ($\nu_2 < -2$) AR(2)-filters, respectively, and the SSA-amplitude functions $\left|D(\nu_i, l)\mathbf{\Gamma}_{AR(2)}(\nu_i) \odot \mathbf{w}\right|$, $i = 1, 2$, of the corresponding SSA designs are below or above the amplitude $|\mathbf{w}|$ of the MSE benchmark towards higher frequencies, assuming an artificial alignment of all amplitude functions at frequency zero for better visual inspection.

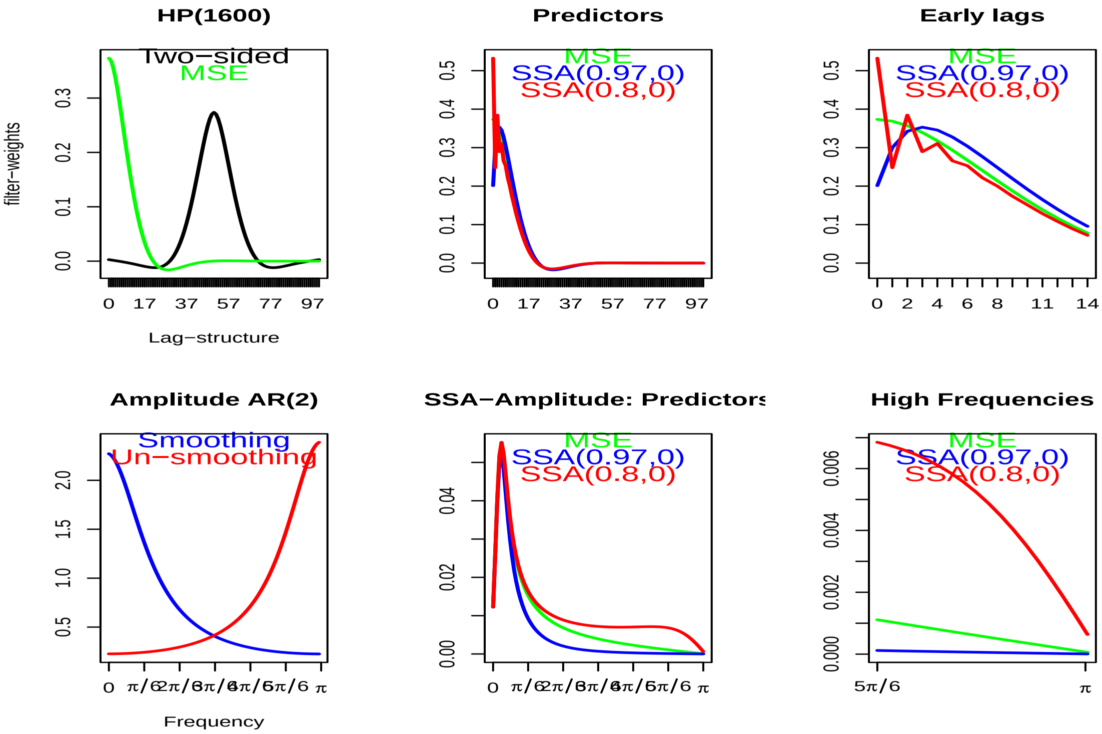

Figure 1: HP(1600) and three nowcasts: MSE, SSA(0.97,0) and SSA(0.8,0). Filter coefficients (top graphs) and SSA-amplitude functions (bottom graphs). The first few lags are highlighted in the top rightmost plot. Amplitude of AR(2) (bottom left), of nowcasts (bottom center) and high frequencies (bottom right). All SSA-amplitude functions are artificially aligned at frequency zero.

The proposed frequency domain analysis suggests that a (unit-root) AR(2) based on $|\nu| \leq 2$, cannot be reconciled with the optimal tracking of the target at least for large L. Indeed, while ν is always such that $\mathbf{\Gamma}_{AR(2)}(\nu)$ is well defined under the regularity assumptions of Theorem (1), i.e. $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, ..., L\}$, the corresponding AR(2) transfer (amplitude) function would be subject to an asymptotic singularity at its peak-frequencies $\omega = \pm \arccos(\nu/2)$, as $L \to \infty$, because λ_i are increasingly densely packed in $[-\pi, \pi]$. By assumption (spectral completeness) the spectral weights w_i of $\boldsymbol{\gamma}_\delta$ are non-vanishing, which is invariably the case for classic forecast or signal extraction filters, and therefore the convolution of target and asymptotically unbounded AR(2) would conflict with an optimal approximation of the former by the latter; moreover, the convolution of the AR(2) with $\boldsymbol{\gamma}_0$ would lead to an asymptotically non-stationary nowcast y_t, strongly periodic, if $-2 \leq \nu < 2$, or strongly trending, if $\nu = 2$, in disagreement with the stationary target specification. These findings in the case $|\nu| \leq 2$ are corroborated by a derivation of the SSA solution in the time domain, see the appendix for details. Note that the reference to large filter lengths is required in the above statements because SSA solutions with $|\nu_0| < 2$ might effectively be obtained if L is 'small enough': as an example, if $\rho_1 \approx \rho_{max}(L) = \lambda_1$ in the holding time constraint, then $\nu \approx 2\lambda_1 < 2$ is required for compliance with the constraint. However, this problem could be solved by increasing L or by decreasing ρ_1 and we argue that the corresponding prediction problem is ill posed at the outset in this case: L is disproportionately small as compared to ht_1. In summary, in typical applications the SSA AR(2) filter is a lowpass: $\nu > 2$ for containment of noisy crossings and the conditions imposed by Corollaries (3) and (6) are not limitations. To conclude we briefly compare classic frequency domain and SSA-amplitude functions in Fig. (2).

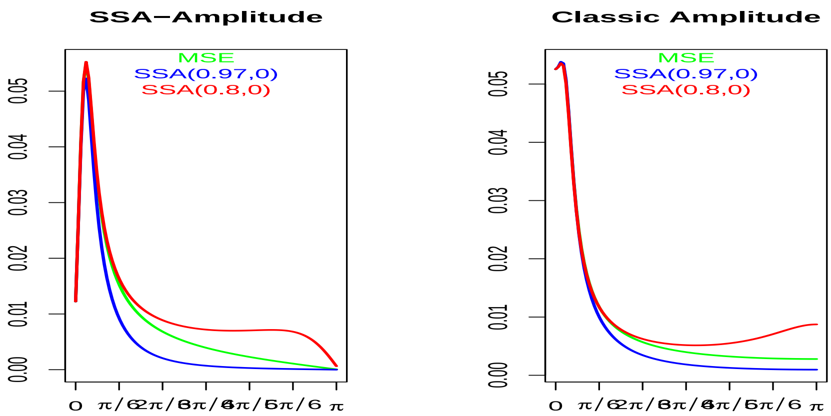

Figure 2: Comparison of SSA-amplitude (left) and classic amplitude functions (right).

The differences are reliant on the choice of the orthonormal basis for the frequency domain decomposition: $(\exp(-ik\omega))_{k=0,...,L-1}$ for the classic approach vs. $(\sin(kj\pi/(L+1)))_{k=1,...,L}$ for SSA, noting that the latter basis ensures compliance with the boundary constraints $b_{-1}(\nu) = b_L(\nu) = 0$ since $\sin(kj\pi/(L+1)) = 0$ for $k \in \{0, L+1\}$. In any case, high-frequency noise handling by SSA remains qualitatively unaffected by the basis whose choice mainly affects the graphical interface for understanding, explaining or interpreting the solution to Criterion (1). In particular, the exact convolution result $\mathbf{w} \odot \mathbf{\Gamma}_{AR(2)}(\nu)$ that helps explain the action of the filter does not hold for the classic basis $(\exp(-ik\omega))_{k=0,...,L-1}$.

4.2 MSE, Zero Crossings and Sign Accuracy

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

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

so that maximization of the target correlation $\rho(y, z, \delta)$ or of SA are equivalent optimization principles. The link to SA motivated the choice of the objective function of Criterion (2) in the first place, see Wildi (2024), and additional justification will be provided shortly hereafter. Proceeding further, we consider the so-called *holding time* (ht) defined by $ht(y|\mathbf{b}, i) := E[t_i - t_{i-1}]$, where t_i, $i \geq 1$ are *consecutive* zero-crossings of y_t, i.e., $t_{i-1} < t_i$, $t_1 \geq L$, $\text{sign}(y_{t_i-1} y_{t_i}) < 0$ for all i and $\text{sign}(y_{t-1} y_t) > 0$ if $t_{i-1} < t < t_i$. Under the above stationarity assumptions, $ht(y|\mathbf{b}, i) = ht(y|\mathbf{b})$ does not depend on i and is the expected duration between consecutive zero-crossings of y_t, see Kedem (1986).

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

$$ht(y|\mathbf{b}) = \frac{\pi}{\arccos(\rho(y, y, 1))}, \tag{15}$$

We refer to Kedem (1986) for proof. The bijective link between the holding time and the lag-one autocorrelation in Equation (15) suggests that Criterion (1) can be interpreted as a maximization of SA under a fixed expected rate of zero crossings of the predictor[3]. In its dual form, the predictor generates the fewest crossings in the long term for a given tracking accuracy, see Corollary (6). Table (1) compares target correlations, sign accuracy, lag-one ACF and holding times of the filters in the previous section. A comparison of holding times of target and MSE predictor in

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

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

the first two columns suggests that the latter is subject to substantial leakage. Indeed, unwanted 'noisy' crossings are often clustered in the vicinity of target crossings, when both filters hover over the zero line. We then argue that an explicit control of noisy crossings due to an unduly small holding time of the (classic MSE) predictor, is a relevant objective, see Wildi (2024) for an application to real time business cycle analysis. Moreover, Criterion (1) ensures an optimal tracking of the target by SSA: this property warrants that the interpretation or the economic content supported by z_t, such as, e.g., a business cycle indicator, can be transferred to SSA. Finally, SSA effectively minimizes, if $\nu > 2$, or maximizes, if $\nu < -2$, the rate of zero-crossings in the class of all predictors with the same target correlation (dual interpretation) and therefore Criterion (1) addresses the problem of noisy false alarms in some way optimally. Since zero-crossings, sign accuracy and correlations are indifferent to the scaling of the filter, we can select $s := \sum_{k=0}^{L-1} \gamma_{k+\delta} b_k / \sum_{k=0}^{L-1} b_k^2 = \sum_{k=0}^{L-1} \gamma_{k+\delta} b_k$ such that MSE performances are optimized by $s\mathbf{b}$, conditional on the imposed holding time constraint. In fact, Criterion 1 could be formulated in terms of

$$\left.\begin{array}{c} \min_{\mathbf{b}}(\boldsymbol{\gamma}_\delta - \mathbf{b})'(\boldsymbol{\gamma}_\delta - \mathbf{b}) \\ \mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1 \\ \mathbf{b}'\mathbf{b} = l \end{array}\right\} \tag{16}$$

where now the objective function $(\mathbf{b} - \boldsymbol{\gamma}_\delta)'(\mathbf{b} - \boldsymbol{\gamma}_\delta)$ is the MSE and l is a free length parameter. The corresponding Lagrangian heads to a system of equations for $\mathbf{b}$

$$\begin{array}{rcl} 2(\boldsymbol{\gamma}_\delta - \mathbf{b}) & = & 2\tilde{\lambda}_1\mathbf{b} + 2\tilde{\lambda}_2\mathbf{M}\mathbf{b} \\ \boldsymbol{\Psi}^{-1}\mathbf{b} & = & F\boldsymbol{\gamma}_\delta \end{array}$$

The first expression is identical with Equation (38), if $\tilde{\lambda}_1 + 1$ is substituted for $\tilde{\lambda}_1$; in the second expression, $\boldsymbol{\Psi} = (2\mathbf{M} - \psi\mathbf{I})$ with $\psi = -2(\tilde{\lambda}_1 + 1)/\tilde{\lambda}_2$ and $F = 2/\tilde{\lambda}_2$. However, in contrast to the target correlation in Criterion (1), the MSE objective now interferes with the length constraint. Specifically, for given l, the parameter $\psi = \psi(l)$ must be selected to comply with the holding time constraint and the pairing $(l, \psi(l))$ can then be determined such as to maximize the objective function. We infer that Criterion (1) simplifies numerical optimization, by integrating out the

[3]In contrast to classic 'sign fitting' approaches, such as ,e.g., logit models, the SSA criterion fits a linear filter of the data to the proper observations $z_{t+\delta}$. This way, the problem structure is more amenable to numerical optimization and its solution has a smaller variance (increased efficiency) under the posited assumptions.

nuisance parameter l, and that SSA reconciles MSE, sign accuracy and smoothing requirements in a flexible and interpretable way. To conclude, we note that target and predictor can be nearly Gaussian, due to aggregation by the filter (central limit theorem), even if ϵ_t isn't, so that the above transformations, linking correlations, ht and SA, might still be practically relevant despite violation of the Gaussian assumption: an illustrative example is proposed in the appendix and Wildi (2024) shows resilience of the holding time Equation (15) for an application to the S&P-500 Index, whose log-returns conflict overtly with the Gaussian assumption.

4.3 Forecasting: MSE, Sign Accuracy and Holding Time

Consider a one-step-ahead forecast for the MA(2)-process

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

where $\gamma_k = 1, k = 0, 1, 2$, $\delta = 1$ and where ϵ_t are assumed to be known, mainly for simplicity of exposition and to focus on the relevant topics. We compute three different SSA forecast filters $y_{ti}, i = 1, 2, 3$ for z_t: the first two are of identical length $L = 20$ with dissimilar holding times ht =3.74 and 10; the third filter deviates from the second one by selecting $L = 50$; the holding time of the first filter matches the lag-one autocorrelation of z_t and is obtained by inserting $\rho(z, z, 1) = 2/3$ into Equation (15); the second holding time $ht = 10$ is sufficiently different in size to reveal some of the salient features of the approach. In addition, we also consider the MSE forecast $\hat{z}_{t+1}^{MSE} = \epsilon_t + \epsilon_{t-1}$, as obtained by classic time series analysis, as well as a trivial 'lag-by-one' forecast $\hat{z}_{t+1}^{lag\ 1} = z_t$, see Fig. (3) (an arbitrary scaling scheme is applied to SSA filters).

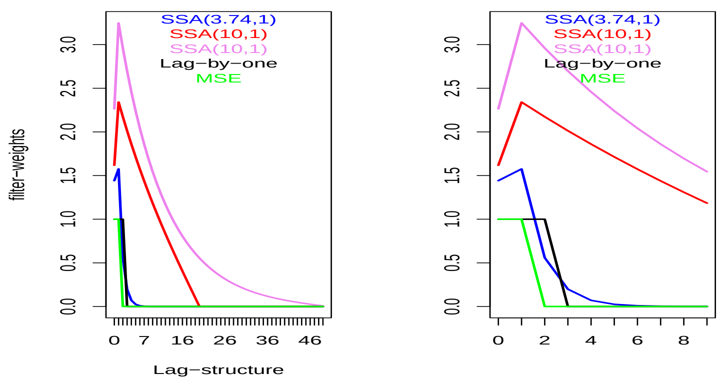

Figure 3: Coefficients of MSE-, SSA- and lag-by-one forecast filters with arbitrarily scaled SSA designs. All lags (left panel) and first ten lags (right panel).

Except for the MSE (green) all other filters rely on past ϵ_{t-k} for $k > q = 2$ which are required for compliance with the holding time constraint (stronger smoothing). For a fixed filter length L, a larger holding time ht asks for a slower zero-decay of filter coefficients (blue vs. red lines) and for fixed holding time ht, a larger L leads to a faster zero-decay but a longer tail of the filter (red vs. violet lines). The distinguishing tips of the SSA filters at lag one in this example are indicative of one of the implicit boundary constraint $b_{-1} = 0$, see Theorem (1). Note that the 'lag-by-one' forecast (black) has the same holding time as the first SSA filter (blue) so that the latter should outperform the former in terms of sign accuracy or, equivalently, in terms of target correlation with the shifted target, as confirmed in Table (2). MSE outperforms all other forecasts in terms of correlation and sign accuracy but it loses in terms of smoothness or holding time; SSA(3.74,0) outperforms the lag-by-one benchmark; both SSA(10,0) loose in terms of sign

	SSA(3.74,1)	SSA(10,1)	SSA(10,1)-long	Lag-by-one	MSE
Target correlation	0.786	0.411	0.427	0.667	0.816
Empirical holding times	3.735	7.998	7.822	3.735	3.000
Empirical sign accuracy	0.788	0.635	0.640	0.732	0.804

Table 2: Performances of MSE and lag-by-one benchmarks vs. SSA: All filters are applied to a sample of length 1000000 of standardized Gaussian noise. Empirical holding times are obtained by dividing the sample length by the number of zero-crossings. The two columns referring to SSA(10,1) correspond to filter lengths 20 (first) and 50 (second).

accuracy but win in terms of smoothness and while the profiles of longer and shorter filters differ in Figure (3), their respective performances are virtually indistinguishable in Table (2), suggesting that the selection of L is to some extent uncritical, assuming it is at least twice the holding time $L \geq 2ht_1$. The table also illustrates the tradeoff between target correlation (or sign accuracy) and holding time, that is formalized by Equation (6). Table (3) allows for a more detailed analysis based on a finer grid of holding time values. In a business cycle application, MSE performances

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

Table 3: Tradeoff: effect of the holding time on target correlation (first row) and sign accuracy (last row) for fixed forecast horizon.

or, equivalently, the target correlation (first row in the table), are related to assessing the level of the cycle, i.e., its precise value above or below the zero-line, whereas the holding time (second row) emphasizes performances at the zero-line, specifically, with the intent to address the number of random crossings due to noise-leakage of one-sided filters. The SSA framework allows to accord the design of the predictor with the particular purpose of the analysis, by a suitable balance of the observed tradeoff. Different filters could be used, in isolation or combination, for measuring the level with higher accuracy, but reduced smoothness, or for assessing sign changes in the growth rate more reliably, see Wildi (2024). In the latter case, given a specified loss in target correlation, the dual reformulation of the SSA criterion in Corollary (6) states that the resulting filter effectively minimizes the rate of zero-crossings or alarms. Formally, a decision for one or several interesting designs could be based on re-computing the above 'tradeoff-table' for the particular prediction problem at hand based on the provided SSA-package. In a final step, we allow the previously fixed forecast horizon $\delta = 1$ to vary, see Fig.(4) for illustration.

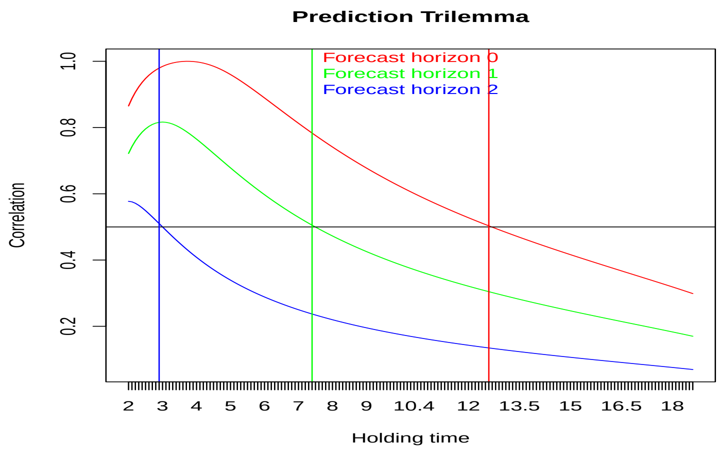

Figure 4: Target correlations of the SSA predictor as a function of the forecast horizon and the holding time.

For each $\delta = 0, 1, 2$, the figure displays the target correlation for given holding times on the abscissa. The peak of the correlation for a given δ appears at a particular holding time value which corresponds to the classic MSE predictor for that forecast horizon: for $\delta = 0$, $y_{t,MSE} = z_t$ and the peak target correlation, obtained at the holding time 3.74 of z_t, is one. To the right of the peak, corresponding to values $\nu > 2$ in Theorem (1), SSA generates fewer crossings (stronger smoothing than MSE) and to its left more ($\nu < -2$); on both sides, SSA maximizes the target correlation, subject to the imposed holding time constraint ht_1 on the abscissa; in its equivalent dual form, SSA maximizes the holding time for given target correlation on the right of the peak ($\nu > 2$), see Corollary (6); finally, the correlation curve for $\delta = 1$ replicates entries in Table (3). For a fixed target correlation, a larger ht_1 (increased smoothness) is functionally related to a smaller δ (reduced timeliness). Specifically, consider the three pairings (ht_{1i}, δ_i), $i = 1, 2, 3$, with values $(2.9, 2)$, $(7.4, 1)$ and $(12.6, 0)$ marked by vertical lines in the figure: the corresponding SSA(ht_{1i}, δ_i)-predictors y_{ti} have a fixed correlation $\rho(y_i, z, \delta_i) = 0.5$ with the target $z_{t+\delta_i}$, marked by the horizontal black line which intersects the curves at the corresponding holding times ht_{1i}, $i = 1, 2, 3$. Fig.(4) generalizes the dilemma in Table (3) to a prediction trilemma, by allowing timeliness, embodied by δ, to become a separate structural element, or hyperparameter, of the estimation problem, together with ht_1. For a particular target $z_{t+\delta_0}$, the pair (ht_1, δ) spans a two-dimensional space of predictors SSA(ht_1, δ) and classic MSE performances can be replicated by selecting $\delta = \delta_0$ and $ht_1 = ht_{MSE}$, the holding time of the MSE predictor. However, alternative priorities in terms of timeliness or smoothness can be triggered by screening the two-dimensional predictor space and our SSA-package can be used to assess an optimal balance of the constituents of the trilemma for general prediction problems, see Wildi (2024) for an application of this framework to business-cycle analysis.

5 Autocorrelation

5.1 Stationary Processes

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

$(\epsilon_t, ..., \epsilon_{t-(L-1)})'$, $\mathbf{b}_\epsilon := \mathbf{\Xi}\mathbf{b}_x$ and consider

$$y_t = \mathbf{b}_x'\mathbf{x}_t \approx (\mathbf{\Xi}\mathbf{b}_x)'\boldsymbol{\epsilon}_t = \mathbf{b}_\epsilon'\boldsymbol{\epsilon}_t, \tag{17}$$

where the approximation by the finite MA inversion of x_t holds if filter coefficients decay to zero sufficiently rapidly (exact results are proposed below). The MSE predictor of $z_{t+\delta}$ is derived in McElroy and Wildi (2020)

$$\hat{\gamma}_{x\delta}(B) = \sum_{k\geq 0}\gamma_{\delta+k}B^k + \sum_{k<0}\gamma_{k+\delta}\left[\boldsymbol{\xi}(B)\right]_{|k|}^{\infty} B^k\boldsymbol{\xi}(B)^{-1}, \tag{18}$$

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

$$\left.\begin{array}{c} \max_{\mathbf{b}_\epsilon} \mathbf{b}_\epsilon'\boldsymbol{\gamma}_{\Xi\delta} \\ \mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon = \rho_1 \\ \mathbf{b}_\epsilon'\mathbf{b}_\epsilon = 1 \end{array}\right\} \tag{19}$$

The SSA solution $\mathbf{b}_x = \mathbf{\Xi}^{-1}\mathbf{b}_\epsilon$ is obtained by solving for $\mathbf{b}_\epsilon$ in Corollary (3), inserting $\boldsymbol{\gamma}_{\Xi\delta}$ for $\boldsymbol{\gamma}_\delta$ in Equation (4). If y_t is (nearly) Gaussian, then $ht_1 := \pi/\arccos(\rho_1)$ expresses the holding time of the predictor and the dual interpretation in Corollary (6) applies invariably. We now derive an exact expression for the SSA predictor. In order to simplify notation we assume an AR(1)-process for $x_t = a_1 x_{t-1} + \epsilon_t$ but our proceeding is otherwise general. In this case $x_{t-L+j} = \sum_{k=0}^{j-1}\xi_k\epsilon_{t-L+j-k} + \xi_j x_{t-L}$ with $\xi_k = a_1^k$ and therefore

$$y_t = \mathbf{b}_x'\mathbf{x}_t = (\mathbf{\Xi}\mathbf{b}_x)'\boldsymbol{\epsilon}_t + \mathbf{b}_x'\boldsymbol{\xi}_L x_{t-L} = \mathbf{b}_\epsilon'\left(\boldsymbol{\epsilon}_t + \mathbf{\Xi}^{-1\,\prime}\boldsymbol{\xi}_L x_{t-L}\right) = \mathbf{b}_\epsilon'\left(\boldsymbol{\epsilon}_t + a_1\mathbf{e}_L x_{t-L}\right), \tag{20}$$

where $\boldsymbol{\xi}_L := (\xi_L, ..., \xi_1)'$, $\mathbf{e}_L = (0, ..., 0, 1)'$ is a unit vector of length L and the last equality follows from definition of $\mathbf{\Xi}^{-1}$ and $\boldsymbol{\xi}_L$. Factually, the approximation error in (17) is coming from replacing $b_{xk}x_{t-k}$ by $b_{xk}(\epsilon_{t-k} + \sum_{j=1}^{L-1-k}\psi_j x_{t-k-j})$ where $x_{t-k} = \epsilon_t + \sum_{j\geq 1}\psi_j x_{t-k-j}$ denotes the AR inversion of x_{t-k}, i.e., the approximation error $b_{xk}\sum_{j\geq 0}\psi_{L-k+j}x_{t-L-j}$ is attributable to those lags in the AR inversion of x_{t-k} which are not part of the filter: for an AR(1), $\psi_j = \left\{\begin{array}{cc} a_1 & j=1 \\ 0 & j>1 \end{array}\right.$ and the rightmost expression in Equation (20) is obtained . We refer to $\mathbf{b}_\epsilon'\boldsymbol{\epsilon}_t$ and $a_1\mathbf{e}_L x_{t-L}$ in terms of main predictor and residual, respectively, and we now consider the exact MSE-predictor $y_{t,\epsilon,MSE}$ as applied to $\boldsymbol{\epsilon}_{t\infty} = (\epsilon_t, \epsilon_{t-1}, ...)'$, the semi infinite extension of $\boldsymbol{\epsilon}_t$, with weights $\boldsymbol{\gamma}_{\Xi\delta\infty} := \mathbf{\Xi}_\infty\hat{\boldsymbol{\gamma}}_{x\delta\infty}$ based on the semi-infinite extension of $\mathbf{\Xi}$ applied to $\hat{\boldsymbol{\gamma}}_{x\delta\infty}$ specified by Equation (18). For a derivation of the exact SSA solution we aim at maximizing the target correlation of y_t with $y_{t,\epsilon,MSE}$ subject to exact length and holding time constraints.

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

$$\mathbf{b}_\epsilon(\nu_1) \quad = \quad D(\nu_1, l)\tilde{\boldsymbol{\nu}}^{-1}\left(\boldsymbol{\gamma}_{\Xi\delta 1:L} + a_1\mathbf{e}_L\boldsymbol{\xi}_\infty'\boldsymbol{\gamma}_{\Xi\delta(L+1):\infty}\right), \tag{21}$$

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

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

and the pairing $(\nu_1, D(\nu_1, l))$ ensures compliance with holding time and length constraints.

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

$$\mathbf{b}_\epsilon' \left(\mathbf{I} + \frac{a_1^2}{1-a_1^2}\mathbf{e}_L\mathbf{e}_L'\right)\mathbf{b}_\epsilon = 1, \tag{22}$$

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

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

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

$$\mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon + a_1\mathbf{b}_\epsilon'\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon + \frac{a_1^3}{1-a_1^2}\mathbf{b}_\epsilon\mathbf{e}_L\mathbf{e}_L'\mathbf{b}_\epsilon, \tag{23}$$

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

$$E[y_t y_{t,\epsilon,MSE}] = \mathbf{b}_\epsilon'\left(\boldsymbol{\gamma}_{\Xi\delta 1:L} + a_1\mathbf{e}_L\boldsymbol{\xi}_\infty'\boldsymbol{\gamma}_{\Xi\delta(L+1):\infty}\right) \tag{24}$$

for the covariance of y_t and $y_{t,\epsilon,MSE}$. If the length constraint (22) is imposed, then the covariance (24) is proportional to the target correlation and therefore we can proceed to optimization, maximizing (24) subject to (22) and (23). Taking derivatives of the Lagrangian then leads to a system of equations for $\mathbf{b}_\epsilon$

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

from which Equation (21) can be inferred. The resulting solution maximizes the correlation with the exact (infinite length) MSE-predictor, up to a fixed scaling constant and subject to exact lag-one and length constraints and therefore $y_t = \mathbf{b}_x'\mathbf{x}_t = \mathbf{b}_\epsilon'\left(\boldsymbol{\epsilon}_t + a_1\mathbf{e}_L x_{t-L}\right)$ can be interpreted as (exact) SSA predictor, as claimed. □

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

To simplify exposition, we now assume pertinence of the finite MA approximation of Equation (17), referring to Proposition (3) for a derivation of exact solutions. For illustration, the generalized Criterion (19) is now applied to the HP target in the previous section, relying on three different AR(1) processes $x_t = a_1x_{t-1} + \epsilon_t$ with $a_1 = -0.6, 0, 0.6$, see Figure (5).

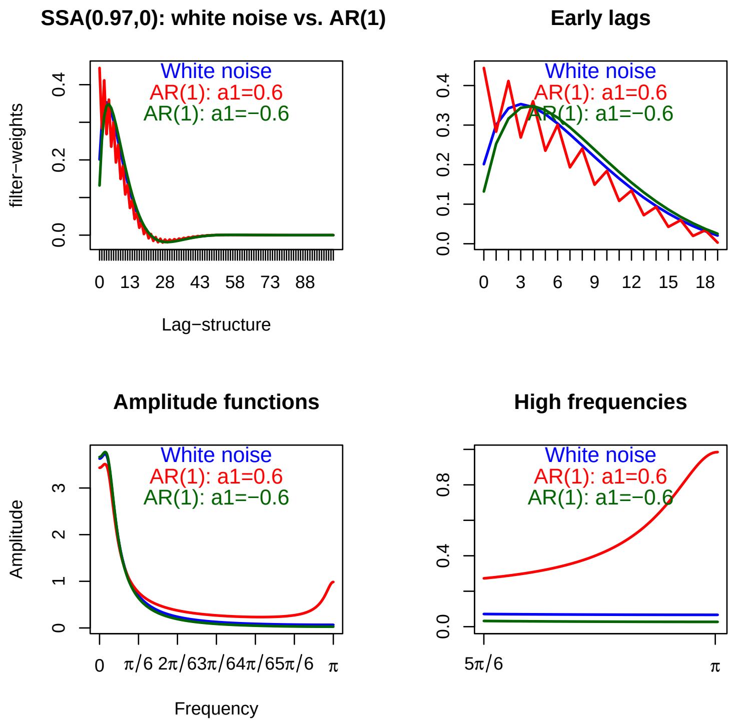

Figure 5: SSA(0.97,0) based on HP(1600)-target. Top left: filters applied to white noise (blue) and AR(1) (red and green); top-right: early lags; bottom-left: classic amplitude functions; bottom-right: classic amplitude towards higher frequencies. All filters are arbitrarily scaled to unit length.

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

Table 4: Holding times of HP (MSE predictor) and SSA as applied to three different AR(1) processes. SSA maintains a fixed holding time across processes.

Table (4) reports holding times of MSE and SSA predictors: while the former depend on the data generating process (DGP), increasing markedly with a_1, the latter remain fixed, irrespective of the DGP. We deduce that the application of a fixed filter to data with unequal dependence structure can lead to qualitatively different components, for example trends or cycles in the case of HP, and SSA can address that ambiguity. For the first two processes in the first two columns of Table (4), the holding times of HP are smaller than the SSA-specification $ht = 12.79$ and SSA must increase smoothness over the benchmark. In contrast, the holding time $ht = 14.74$ of the benchmark for the third process exceeds the SSA-specification and the latter is asked to generate additional noisy crossings over the benchmark. This atypical demand is reflected by the ripples of the corresponding filter coefficients in Fig.(5). In the frequency domain, the tail behavior of the (classic) amplitude function marks control of the rate of zero-crossings: for $a_1 = -0.6$ the filter damps high-frequency noise most effectively; for $a_1 = 0.6$ increased leakage towards frequency π permits the generation of excess noisy crossings while maintaining optimal tracking of the target by the filter.

5.2 Integrated Processes

As shown, the main modification of the original Criterion (1) in the case of stationary processes concerns the target specification based on the MSE predictor in Equation (18). We now consider an extension to non-stationary integrated processes: let $\tilde{x}_t$ be such that $\Delta(B)\tilde{x}_t = (1-B)^d\tilde{x}_t$ is stationary and invertible and assume $\sum_{k=-\infty}^{\infty}|\gamma_k k^d| < \infty$. Then the MSE predictor is

$$\hat{\gamma}_{\tilde{x}\delta}(B) = \sum_{k\geq 0}\gamma_{\delta+k}B^k + \sum_{k<0}\gamma_{k+\delta}\left(\sum_{j=1}^{d}A_{j,d-k}B^{d-j} + \sum_{j=1}^{|k|}\psi_{-k-j}\left[\boldsymbol{\xi}(B)\right]_{|k|}^{\infty}B^{-j}\Delta(B)\boldsymbol{\xi}(B)^{-1}\right), \quad (25)$$

where $(1-B)^d =: 1-\sum_{j=1}^{d}\delta_j B^j$, $\Psi(B) := 1/\Delta(B) = \sum_{j\geq 0}\psi_k B^k$ and $A_{jt} := \psi_{t-j} - \sum_{k=1}^{d-j}\delta_k\psi_{t-j-k}$, see McElroy and Wildi (2020). While the general intuition behind the MSE predictor remains the same as for the stationary case, the additional term in $A_{j,d-k}$ accounts for a polynomial $p(t)$ of t in the null space of the difference operator $\Delta(B)$, i.e., a solution to the homogeneous difference equation $\Delta(B)p(t) = 0$: the coefficients of this polynomial are determined by a proper initialization of the process (boundary restrictions with permanent effect). We then infer that the MSE predictor can be decomposed into 'pure' MA inversion and $p(t)$, the latter for matching boundary conditions, recall the proof of Proposition (4). For a derivation of the SSA solution we generally discard $p(t)$ which can be cancelled by imposing suitable structure or constraints to the predictor, see below for details. For further discussion it is now convenient to assume $d = 1$ so that $\Delta(B) = 1 - B$ and the solution to the homogeneous equation $p(t) = x_0$ is a constant which is assumed to be determined by the initialization x_0 at $t = 0$. Consider the following notation: $\boldsymbol{\Sigma}$ is the $L \cdot L$ dimensional summation matrix, with ones along and below the main diagonal and $\boldsymbol{\Delta} := \boldsymbol{\Sigma}^{-1}$, $\tilde{\boldsymbol{\Xi}} := \boldsymbol{\Sigma}\boldsymbol{\Xi}$; $\hat{\boldsymbol{\gamma}}_{\tilde{x}\delta}$ denotes the first L coefficients of the MSE predictor (25), ignoring $p(t) = x_0$, and $\boldsymbol{\gamma}_{\tilde{\Xi}\delta} := \tilde{\boldsymbol{\Xi}}\hat{\boldsymbol{\gamma}}_{\tilde{x}\delta}$ so that $y_{MSE,t} \approx \hat{\boldsymbol{\gamma}}'_{\tilde{x}\delta}\mathbf{x}_t \approx \boldsymbol{\gamma}'_{\tilde{\Xi}\delta}\boldsymbol{\epsilon}_t$; finally, let $\mathbf{b}_{\tilde{\epsilon}} := \tilde{\boldsymbol{\Xi}}\mathbf{b}_x$, so that

$$y_t = \mathbf{b}'_x\mathbf{x}_t = \mathbf{b}'_x\boldsymbol{\Sigma}'\boldsymbol{\Delta}'\mathbf{x}_t \approx \mathbf{b}'_x\boldsymbol{\Sigma}'\boldsymbol{\Xi}'\boldsymbol{\epsilon}_t = \mathbf{b}_{\tilde{\epsilon}}\boldsymbol{\epsilon}_t,$$

where we used $\boldsymbol{\Sigma}\boldsymbol{\Xi} = \boldsymbol{\Xi}\boldsymbol{\Sigma}$. All approximations hold if filter coefficients decay sufficiently rapidly towards zero in absolute value and we can eventually ameliorate the fit by shifting the data by $-x_0$ to eliminate $p(t)$. Note, however, that z_t and hence y_t are generally non-stationary[4] so that lag-one ACF and the holding time are not properly defined anymore. Therefore, we propose to emphasize constraints in first differences, making use of the identity $\mathbf{b}'_{\tilde{\epsilon}}\boldsymbol{\gamma}_{\tilde{\Xi}\delta} = \mathbf{b}'_{\epsilon}\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$ for the objective function, where $\mathbf{b}_{\epsilon} := \boldsymbol{\Xi}\mathbf{b}_x = \boldsymbol{\Delta}\mathbf{b}_{\tilde{\epsilon}}$. We then obtain two equivalent expressions for the generalized SSA criterion

$$\left.\begin{array}{c}\max_{\mathbf{b}_\epsilon}\mathbf{b}'_\epsilon\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}\\ \mathbf{b}'_\epsilon\mathbf{M}\mathbf{b}_\epsilon = \rho_1 l\\ \mathbf{b}'_\epsilon\mathbf{b}_\epsilon = l\\ \mathbf{b}'_\epsilon\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_\epsilon = 1\end{array}\right\} \quad \text{or} \quad \left.\begin{array}{c}\min_{\mathbf{b}_{\tilde{\epsilon}}}(\boldsymbol{\gamma}_{\tilde{\Xi}\delta} - \boldsymbol{\Sigma}\mathbf{b}_\epsilon)'(\boldsymbol{\gamma}_{\tilde{\Xi}\delta} - \boldsymbol{\Sigma}\mathbf{b}_\epsilon)\\ \mathbf{b}'_\epsilon\mathbf{M}\mathbf{b}_\epsilon = \rho_1 l\\ \mathbf{b}'_\epsilon\mathbf{b}_\epsilon = l\end{array}\right\} \quad (26)$$

The criterion on the left emphasizes the target correlation and optimizes $\mathbf{b}_\epsilon$ subject to two length constraints: the first one $\mathbf{b}'_\epsilon\mathbf{b}_\epsilon = l$ ensures that ρ_1 is the lag one ACF of $\mathbf{b}_\epsilon{}'\boldsymbol{\epsilon}_t \approx y_t - y_{t-1}$; the second one $\mathbf{b}'_\epsilon\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_\epsilon = 1$ warrants proportionality of objective function and target correlation. The second criterion on the right emphasizes a MSE objective and therefore the additional length constraint $\mathbf{b}'_\epsilon\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_\epsilon = 1$ can be skipped. In both cases, l is an additional free parameter and can be selected to maximize the respective objective, recall the discussion following Criterion (16). For the left hand optimization, the derivative of the Lagrangian heads towards a system of equations for $\mathbf{b}_\epsilon$

$$\boldsymbol{\mathcal{V}}^{-1}\mathbf{b}_\epsilon = D\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta} \quad (27)$$

[4]The target z_t can be stationary if γ_k cancels the unit root(s) of x_t such as for example in the case of a bandpass or highpass target filter.

with $\boldsymbol{\mathcal{V}} := 2\mathbf{M} - \nu_1\mathbf{I} - \nu_2\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$ and for given ν_2, $\nu_1 = \nu_1(\nu_2)$ ensures compliance with the holding time constraint. We can then select ν_2 such that a solution to the holding time constraint exists (Theorem (1) asserts that $\nu_2 = 0$ is a possible choice) and that the objective function is maximized. A similar layout applies to the right hand criterion, for which the derivative of the objective function becomes $-2\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta} + 2\boldsymbol{\Sigma}'\boldsymbol{\Sigma}\mathbf{b}_\epsilon$ leading to the Lagrangian equations $\left(-\tilde{\lambda}_2\mathbf{M} - \tilde{\lambda}_1\mathbf{I} + \boldsymbol{\Sigma}'\boldsymbol{\Sigma}\right)\mathbf{b}_\epsilon = \boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$ such that

$$\tilde{\boldsymbol{\mathcal{V}}}^{-1}\mathbf{b}_\epsilon = F\boldsymbol{\Sigma}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta}$$

with $\tilde{\boldsymbol{\mathcal{V}}} := 2\mathbf{M} - \nu_1\mathbf{I} + F\boldsymbol{\Sigma}'\boldsymbol{\Sigma}$. We can now select $F, \nu_1(F)$ such that the holding time constraint applies and the objective is maximized. The main difference with Equation (27) is that the optimal F is also the optimal MSE scaling. Note that since both criteria rely on the determination of two parameters, there is no preference for the target correlation (left criterion) over MSE (right criterion) as was previously the case for stationary processes: both SSA optimization principles are deemed equal. For a (nearly) Gaussian predictor y_t, $ht_1 := \pi/\arccos(\rho_1)$ expresses the holding time of $y_t - y_{t-1}$: interpreted in its dual form, y_t is 'most monotonic' in the sense that sign changes of $y_t - y_{t-1}$ are fewest possible for a given tracking accuracy.

In a next step, we propose to enhance the predictor by adding structure for tracking the (asymptotically unbounded) level of the non-stationary target. Consider $E[z_t|x_0] = \sum_{|k|<\infty}\gamma_k x_0 = \Gamma(0)x_0$ and $E[y_t|x_0] = \sum_{k=0}^{L-1} b_{xk}x_0 = \hat{\Gamma}_x(0)x_0$, where $\Gamma(\omega), \hat{\Gamma}_x(\omega)$ denote the transfer functions (z-transforms evaluated on the unit circle $z = \exp(i\omega)$) of the sequences $\boldsymbol{\gamma}$ and $\mathbf{b}_x$, respectively. We can impose a vanishing bias by requiring $\Gamma(0) = \hat{\Gamma}_x(0)$: in this case $p(t) = x_0$ is cancelled in relative terms and can be skipped from the MSE predictor in the objective function without affecting optimization, as claimed. Moreover, applying a first order Taylor approximation to (the transfer function of) the prediction error filter $\Gamma(\omega) - \hat{\Gamma}_x(\omega)$ centered at $\omega = 0$, using $\sum_{|k|<\infty}|\gamma_k k| < \infty$ for computing the derivative, shows that the error filter cancels the unit root of x_t such that the prediction error $z_{t+\delta} - y_t$ is stationary (predictor and target are cointegrated). Clearly, these properties are desirable in the present context[5] and therefore we now impose the zero-bias (cointegration) constraint $\Gamma(0) = \hat{\Gamma}_x(0)$ which can be expressed in vector notation as $\mathbf{b}_x = \Gamma(0)\mathbf{e}_1 - \mathbf{B}\tilde{\mathbf{b}}$, where $\mathbf{B}$ is an $L \cdot (L-1)$ dimensional matrix, whose first row, filled with -1, is stacked on the $(L-1)\cdot(L-1)$identity, and where the unit vector $\mathbf{e}_1 = (1, 0, ..., 0)'$ and $\tilde{\mathbf{b}} = (b_1, ..., b_L)'$ are of length L and $L-1$, respectively. The holding-time constraint of (either) Criterion (26) then becomes

$$\rho_1 l = \mathbf{b}_\epsilon'\mathbf{M}\mathbf{b}_\epsilon = \mathbf{b}_x'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{b}_x = \Gamma(0)^2\mathbf{e}_1'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{e}_1 - 2\Gamma(0)\mathbf{e}_1'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{B}\tilde{\mathbf{b}} + \tilde{\mathbf{b}}'\mathbf{B}'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{B}\tilde{\mathbf{b}}$$

Taking derivatives with respect to $\tilde{\mathbf{b}}$ gives $-2\Gamma(0)\mathbf{e}_1'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{B} + 2\mathbf{B}'\boldsymbol{\Xi}'\mathbf{M}\boldsymbol{\Xi}\mathbf{B}\tilde{\mathbf{b}}$; the derivative of the length constraint $\mathbf{b}_\epsilon'\mathbf{b}_\epsilon = l$ is obtained by inserting $\mathbf{I}$ for $\mathbf{M}$ in the previous expression. Finally, for the objective function we here consider the right hand (MSE-) SSA criterion:

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

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

$$\tilde{\mathbf{b}} = \left(\mathbf{B}'\boldsymbol{\Xi}'\boldsymbol{\nu}\boldsymbol{\Xi}\mathbf{B} - D\mathbf{B}'\tilde{\boldsymbol{\Xi}}'\tilde{\boldsymbol{\Xi}}\mathbf{B}\right)^{-1}\left\{D\left(\mathbf{B}'\tilde{\boldsymbol{\Xi}}'\boldsymbol{\gamma}_{\tilde{\Xi}\delta} - \Gamma(0)\mathbf{B}'\tilde{\boldsymbol{\Xi}}'\tilde{\boldsymbol{\Xi}}\mathbf{e}_1\right) + \Gamma(0)\mathbf{B}'\boldsymbol{\Xi}'\boldsymbol{\nu}\boldsymbol{\Xi}\mathbf{e}_1\right\} \tag{28}$$

with $\boldsymbol{\nu} = 2\mathbf{M} - \nu\mathbf{I}$, $D = 1/\tilde{\lambda}_2$ and $\nu = -2\tilde{\lambda}_1/\tilde{\lambda}_2$, where $\tilde{\lambda}_1, \tilde{\lambda}_2$ are the Lagrange multipliers of length and holding time constraints, respectively. The solution to the right hand (MSE) Criterion (26), reparameterized in terms of $\tilde{\mathbf{b}}$, is obtained from Equation (28) whereby the pairing

[5] The constraint could be imposed in the stationary case, too, but the benefit of improved level tracking is more explicit for integrated processes with asymptotically unbounded paths.

$(\tilde{\lambda}_1, \tilde{\lambda}_2)$ or, equivalently, $(l, \nu(l))$ must solve the holding time constraint and minimize MSE. Finally, extensions to higher order integration orders $d > 1$ can be obtained by assuming γ_k to be such that $\sum_{|k|<\infty} |\gamma_k k^d| < \infty$ and by imposing additional (cointegration) constraints of the form $\sum_{k=0}^{L-1} b_k k^j = \sum_{k=-\infty}^{\infty} \gamma_k k^j,\ j = 0, ..., d-1$, cancelling the higher order polynomial $p(t)$, i.e., the terms in $A_{j,d-k} B^{d-j}$ of Equation (25), and ensuring stationarity of the prediction error[6]. By imposing additional structure to the filter weights of the SSA predictor, the solution to the homogeneous difference equation is cancelled and the objective function can emphasize the pure MA inversion of the MSE predictor in Equation (25).

6 Smoothing: SSA vs. Whittaker Henderson and Hodrick Prescott

We here consider Criterion (1) without explicit connection to model assumptions about the data so that the objective function is not related to MSE performances or prediction anymore. An interesting simple SSA smoothing-problem is obtained when selecting $\boldsymbol{\gamma} = 1$ the identity and $\delta = -(L-1)/2$ (backcast), assuming L to be an odd number. In this case, the solution y_t to

$$\left.\begin{array}{c} \max_{\mathbf{b}} \rho\Big(y, x, \delta = -(L-1)/2\Big) \\ \rho(y, y, 1) = \rho_1 \end{array}\right\} \tag{29}$$

aims at tracking $x_{t+\delta}$ while being smoother if $\rho_1 > \rho(x, x, 1)$, the lag-one ACF of the data. Selecting $\delta = -(L-1)/2$ ensures symmetry of the backcast: the coefficients of the causal filter are centered at $x_{t-(L-1)/2}$ and the tails are mirrored at the center point. Let then $\tilde{y}_t := y_{t-\delta}$ denote the so-called SSA smoother, the solution of Criterion (29) shifted forward and centered at x_t. The proposed optimization problem can be considered as an original smoothing algorithm, since the identity target $\boldsymbol{\gamma} = 1$ does not refer to a specific benchmark anymore. We can then contrast our approach with classic Whittaker-Henderson (WH) smoothing, see Whittaker (1922), who proposes to solve the following optimization problem for $\mathbf{u} := (u_1, ..., u_T)$

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

The HP filter is obtained by selecting $d = 2$, emphasizing squared second-order differences, i.e., the curvature, in the penalty term. In the case of stationary data, increasing λ typically leads to a longer holding time of u_t but Criterion (29) is more apt at controlling this particular characteristic. For illustration, Fig.(6) displays HP and two different SSA smoothers, all based on a fixed length $L = 201$, and Table (5) compares their performances. For an identical holding time, SSA1 (blue

	HP	SSA1	SSA2
Holding times	34.366	34.366	42.830
Target correlations	0.271	0.301	0.271
RMS second-order differences	0.204	0.757	0.547

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

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

[6] One can use a d-th order Taylor approximation of the error filter $\Gamma(\omega) - \hat{\Gamma}_x(\omega)$ centered at $\omega = 0$ to show that the error filter cancels the d-th order unit root of x_t.

Eventually, a hybridization obtained by plugging SSA on WH, such as discussed in the previous section, could bridge the gap between both smoothing concepts.

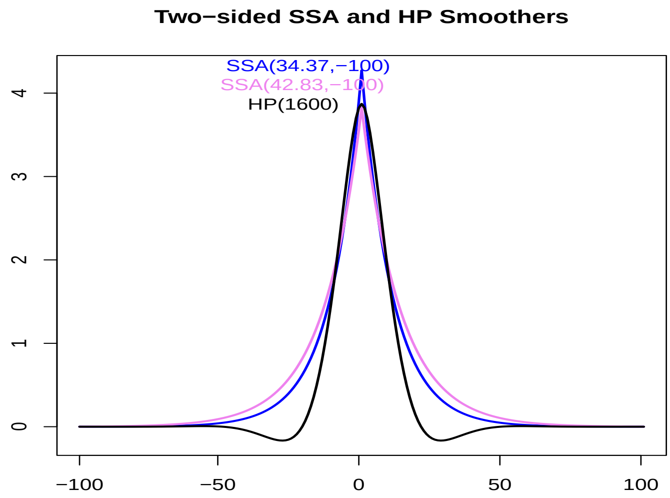

Figure 6: Coefficients of two-sided SSA and HP smoothers of length 401, arbitrarily scaled to unit length: the first SSA design (blue line) replicates the holding time of HP, the second SSA design (violet line) replicates the tracking-ability or target correlation of HP.

To conclude, we broaden the scope of the above comparison. An optimal causal or one-sided SSA smoother can be obtained straightforwardly, by specifying $\delta = 0$, instead of $\delta = -(L-1)/2$, in Criterion (29). Furthermore, the original Criterion (1) or its extension (26) address more general estimation or prediction problems than the simple identity-smoothing (29), based on $\boldsymbol{\gamma} = 1$. Finally, SSA can accommodate for the DGP in terms of the MA inversion of x_t. We argue that if a predictor, or a smoother, has to match sign-changes of a target while keeping control of the alarm rate, determined by zero-crossings, as well as of MSE performances and timeliness characteristics, then the hyperparameters ρ_1, δ of SSA, as well as the underlying optimization principle, address facets of that problem in a more nuanced way than λ for the WH smoother and, incidentally, for the HP filter.

7 Appendix

7.1 Theoretical vs. Empirical Holding Times for t-Distributed Random Variables

Table (6) evaluates the effect of heavy tails on holding times of the SSA nowcasts of Section (3). Heavier tails increase the positive holding time bias because extreme observations can trigger the impulse response, which does not change sign frequently. On the other hand, the central limit effect works against this bias in the sense that stronger smoothing of the non-Gaussian noise by the filter can narrow the gap separating the predictor from Gaussianity: as an example, the filter in the second column seems least affected by distortions of the holding time, in relative terms. In any case, $ht = ht_{\mathbf{b}} := \pi / \arccos(\rho(y, y, 1)$ is named holding time: if y_t is (nearly) Gaussian, then the mean duration between consecutive zero crossings converges to ht for long samples. An extension to (conditional) heteroscedastic processes is discussed in Wildi (2024) who illustrates

	MSE	SSA(0.97,-100)	SSA(0.8,-100)
t-dist.: df=2.1	9.89	14.14	5.98
t-dist.: df=4	8.87	13.34	5.32
t-dist.: df=6	8.55	13.10	5.14
t-dist.: df=8	8.44	13.01	5.04
t-dist.: df=10	8.30	12.89	4.98
t-dist.: df=100	8.16	12.83	4.90
Gaussian	8.14	12.79	4.88

Table 6: The effect of heavy tails on the empirical holding times of HP predictors, based on samples of length one Million: Gaussian vs. t-distributed data.

that SA and ht are fairly robust against vola-clustering.

7.2 Spherical Length- and Hyperbolic Holding-Time Constraints

Consider the spectral decomposition

$$\mathbf{b} := \sum_{i=1}^{L} \alpha_i \mathbf{v}_i \tag{30}$$

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

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

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

$$\alpha_{i_0} = \pm\sqrt{\frac{\lambda_{j_0} - \rho_1}{\lambda_{j_0} - \lambda_{i_0}} - \sum_{k\neq i_0, k\neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}}} \tag{31}$$

Consider the case $\rho_1 = -\rho_{max}(L) = \lambda_L$ and set $i_0 = L$:

$$\alpha_L = \pm\sqrt{1 - \sum_{k\neq L, k\neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_L}} \tag{32}$$

For $j_0 = L-1$ we have $\lambda_{L-1} - \lambda_k < 0$ in the nominators of the summands of 32 and $\lambda_{L-1} - \lambda_L > 0$ in the denominators. We then deduce that if $\alpha_k \neq 0$ for some $k < L-1$, then $|\alpha_L| > 1$ which would contradict the unit-sphere constraint. Therefore, $\alpha_k = 0$ for $k < L-1$ so that $\alpha_L = \pm 1$, $\alpha_{L-1} = 0$ and $\mathbf{b} := \pm \mathbf{v}_L$ (the contacts of unit-sphere and hyperbola are tangential at the vertices $\pm \mathbf{v}_L$). Since $w_L \neq 0$ (completeness assumption), the SSA solution $\mathbf{b} := \text{sign}(w_L)\mathbf{v}_L$ warrants a positive objective function $\boldsymbol{\gamma}_\delta' \mathbf{b} = \text{sign}(w_L) w_L > 0$, confirming Corollary (1). Next, for $\rho_1 > \lambda_L$ the quotient $\frac{\lambda_{L-1}-\rho_1}{\lambda_{L-1}-\lambda_L}$ in (31) is smaller one which allows for non-vanishing $\alpha_k \neq 0$, $k < L-1$. However, in this case the number under the square root should remain positive which is always the case for $\rho_1 \leq \rho_{max}(L) = \lambda_1$ since the term $-\alpha_1^2 \frac{\lambda_{L-1}-\lambda_1}{\lambda_{L-1}-\lambda_L}$ of the sum $-\sum_{k<L-1} \alpha_k^2 \frac{\lambda_{L-1}-\lambda_k}{\lambda_{L-1}-\lambda_L}$ can compensate for a potentially negative value of $\frac{\lambda_{L-1}-\rho_1}{\lambda_{L-1}-\lambda_L}$. In particular, if $\rho_1 = \rho_{max}(L)$, then positiveness of the term under the square root implies $\alpha_1 = 1$ and $\alpha_2, ..., \alpha_{L-2} = 0$, by symmetry, so that $\alpha_L = \alpha_{L-1} = 0$, i.e., $\mathbf{b} = \pm v_1$, confirming again Corollary (1). In between, that is for $\rho_{min}(L) < \rho_1 < \rho_{max}(L)$, the number under the square root is in the open unit intervall $]0, 1[$ and the intersection of unit sphere and holding time hyperbola is non-empty and of dimension $L - 2 \geq 1$.

7.3 Time Domain SSA Solution

A time domain SSA solution is obtained from the difference equations (40), assuming the stable boundary conditions $b_{-1} = b_L = 0$.

Proposition 4. *Let all regularity assumptions of Theorem* (1) *hold and let* $\mathbf{b}(\nu) = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta$ *with coefficients* $b_k(\nu)$. *Consider the solution* $\tilde{\mathbf{b}}(\nu)$ *of the difference equation*

$$\tilde{b}_{k+1}(\nu) \quad = \quad \nu\tilde{b}_k(\nu) - \tilde{b}_{k-1}(\nu) + D\gamma_{k+\delta} \ , \ 0 \leq k \leq L-1 \tag{33}$$

with arbitrary initializations $\tilde{b}_0(\nu), \tilde{b}_1(\nu)$. *If* $\nu \in]-2, 2[$, *then* $\mathbf{b}(\nu)$ *is given by*

$$b_k(\nu) = \tilde{b}_k(\nu) + C(\nu)\cos(\phi(\nu) + \arccos(\nu/2)k) \tag{34}$$

where $C(\nu), \phi(\nu)$ *can be chosen such that* $b_{-1}(\nu) = b_L(\nu) = 0$. *If* $\nu = 2$ *then*

$$b_k(\nu) = \tilde{b}_k(\nu) + c(\nu) + a(\nu)k \tag{35}$$

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

$$b_k(\nu) = \tilde{b}_k(\nu) + c(\nu) + a(\nu)(-1)^k k \tag{36}$$

where $a(\nu), c(\nu)$ *can be selected such that* $b_{-1}(\nu) = b_L(\nu) = 0$. *Finally, if* $|\nu| > 2$ *then*

$$b_k(\nu) = \tilde{b}_k(\nu) + C_1(\nu)\lambda^k + C_2(\nu)\lambda^{L-k} \tag{37}$$

where $\lambda \in \mathbb{R} \setminus \{0\}$ *is such that* $\lambda + 1/\lambda = \nu$ *and* $C_1(\nu), C_2(\nu)$ *can be selected such that* $b_{-1}(\nu) = b_L(\nu) = 0$.

Proof: According to Theorem (1) the coefficients $b_k(\nu)$ of $\mathbf{b}(\nu)$ must conform to the same difference equation as $\tilde{\mathbf{b}}(\nu)$, subject to $b_{-1}(\nu) = b_L(\nu) = 0$. If $\tilde{b}_{-1}(\nu) = \tilde{b}_L(\nu) = 0$ then $\mathbf{b}(\nu) = \tilde{\mathbf{b}}(\nu)$ and either $C(\nu) = 0$ in Equation (34) or $a(\nu) = b(\nu) = 0$ in Equation (35) (or (36)) or $C_1 = C_2 = 0$ in Equation (37), depending on ν. Otherwise, the difference $\mathbf{db}(\nu) := \mathbf{b}(\nu) - \tilde{\mathbf{b}}(\nu)$ must be a solution to the homogeneous difference equation

$$db_{k+1}(\nu) - \nu db_k(\nu) + db_{k-1}(\nu) = 0 \ , \ 0 \leq k \leq L-1$$

If $\nu \in]-2, 2[$ then $db_k(\nu) = C(\nu)\cos(\phi(\nu) + \arccos(\nu/2)k)$ where $\phi(\nu), C(\nu)$ can be selected such that $db_{-1}(\nu) = -\tilde{b}_{-1}(\nu)$ and $db_L(\nu) = -\tilde{b}_L(\nu)$, i.e., $b_{-1}(\nu) = b_L(\nu) = 0$. Similarly, for $\nu = 2$ or $\nu = -2$ the solutions to the homogeneous difference equations are $c(\nu) + a(\nu)k$ and $c(\nu) + a(\nu)(-1)^k k$, respectively, and $a(\nu), c(\nu)$ can be selected such that $b_{-1}(\nu) = b_L(\nu) = 0$, as claimed. Finally, if $\nu > 2$, then the roots of the characteristic AR(2) polynomial are real reciprocal $\lambda, 1/\lambda$ with $\lambda + 1/\lambda = \nu$ and the solution to the corresponding homogeneous equation is $C_1(\nu)\lambda^k + C_2(\nu)\lambda^{L-k}$, where $C_1(\nu), C_2(\nu)$ can be selected such that $b_{-1}(\nu) = b_L(\nu) = 0$. $\square$

In a prediction framework, typically, γ_k and b_k should decay towards zero for increasing lag k such that the remote past of a time series becomes increasingly irrelevant for determining its future: we refer to this property as the tacit prediction premise. Also, typically, γ_k decay sufficiently rapidly to ensure convergence of the MA-inversion of the AR(2)-filter in Equation (33).

Proposition 5. *Let all regularity assumptions of Theorem* (1) *hold and assume* $\nu \in [-2, 2]$ *and assume the MA-inversion* $\tilde{b}_k$ *of the AR(2) difference Equation* (33) *is convergent for arbitrary initialization* $\tilde{b}_0, \tilde{b}_1$ *as* $L \to \infty$. *Then the coefficients* $b_k(\nu, L)$ *of* $\mathbf{b}(\nu, L)$ *do not converge towards zero for increasing lag* k, *asymptotically as* $L \to \infty$.

Proof: Let $\tilde{\mathbf{b}}(\nu, L)$ designate the MA inversion of the SSA AR(2) filter, see Equation (33) and assume the initial values $\tilde{b}_0, \tilde{b}_1$ are such that $\tilde{b}_{-1} := \tilde{b}_{-1}(\nu, L) = D\gamma_0 + \nu\tilde{b}_0 - \tilde{b}_1 \neq 0$. If $\nu = 2$ then $c(\nu, L), a(\nu, L)$ in 35 or 36 must converge since they depend on $\tilde{b}_{-1}$ and $\tilde{b}_L(\nu, L)$, noting that $\lim_{L\to\infty} \tilde{b}_L(\nu, L) = 0$, by convergence of the MA inversion. Also, $\lim_{L\to\infty} a(\nu, L) = 0$, by convergence of the MA-inversion, and $c(\nu, L) - a(\nu, L) + \tilde{b}_{-1} = b_{-1}(\nu, L) = 0$ so that $\lim_{L\to\infty} c(\nu, L) = -\tilde{b}_{-1} \neq 0$. Then

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

A similar proof applies to the case $\nu = -2$. If $|\nu| < 2$ is a fixed number, not depending on L, then $C(\nu, L), \phi(\nu, L)$ in 34 must converge (since they depend on $\tilde{b}_{-1}$ and $\tilde{b}_L(\nu, L) \to 0$) and

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

where $\arccos(\nu/2) \neq 0$, by assumption, and $C(\nu, \infty) \neq 0$ since $C(\nu, L)\cos(\phi(\nu, L) - \arccos(\nu/2)) + \tilde{b}_{-1} = b_{-1}(\nu, L) = 0$. Finally, if $\nu = \nu(L) < 2$ is such that $\lim_{L\to\infty} \nu(L) = 2$, then

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

since $C(2, \infty)\cos(\phi(2, \infty))) + \tilde{b}_{-1} = b_{-1}(\nu, L) = 0$.

□

The proposition suggests that the restriction $|\nu| > 2\rho_{max}(L)$ in Corollaries (3) and (6) is not a limitation since the more stringent condition $|\nu| > 2$ is necessary to comply with the tacit prediction premise. Moreover, the unit root solution to the homogeneous AR(2) difference equation, in the case $|\nu| \leq 2$, reflects the asymptotically unbounded peak-value of the transfer function $\mathbf{\Gamma}_{AR(2)}(\nu)$ of the SSA AR(2) filter at the frequency $\arccos(\nu/2)$ as $L \to \infty$, see Section (4.1).

7.4 The Choice of M

The matrix $\mathbf{M}$ is not uniquely determined by the quadratic form $\mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{k=1}^{L-1} b_{k-1}b_k$. Indeed, the family of corresponding matrices $\mathbf{M}(\kappa)$ has zeroes everywhere except above and below the main diagonal, where the corresponding entries are κ and $1-\kappa$, respectively, with $\kappa \in \mathbb{R}$: for $\kappa = 0.5$ the original $\mathbf{M} := \mathbf{M}(0.5)$ is obtained. Evidently, all obtained results straightforwardly extend to $\mathbf{M}(\kappa)$, noting that derivatives of quadratic forms would involve $\mathbf{M}(\kappa) + \mathbf{M}(\kappa)' = 2\mathbf{M}(0.5)$ so that κ is cancelled. More generally, the technical elements of the proof of Theorem (1) assume $\mathbf{M}$ to be symmetric with pairwise different eigenvalues and therefore the theorem would apply to constraints of the ACF at higher lags, though this possibility is neither explicitly required nor explored here.

7.5 Illustration of a Case of Incomplete Spectral Support

In order to illustrate the case of incomplete spectral support addressed by Corollary (2) we consider a simple nowcast example, $\delta = 0$, based on a band-limited target $\boldsymbol{\gamma}_0 = \sum_{i=1}^{7} 0.378\mathbf{v}_i$ of length $L = 10$, where $\mathbf{v}_i$ are the eigenvectors of the $10 \cdot 10$-dimensional $\mathbf{M}$, assuming that the last three weights $w_8 = w_9 = w_{10} = 0$ vanish ($m = 7$ in Equation (3)) and that the first seven weights are constant $w_i = 0.378$: this particular weighting scheme implies that $\boldsymbol{\gamma}_0'\boldsymbol{\gamma}_0 = 1$ such that the SSA objective function is also the target correlation. The left panel in fig. 7 displays the lag-one acf 8 of $\mathbf{b}(\nu)$ given by 7 as a function of $\nu \in [-2, 2] - \{2\lambda_i, i = 1, ..., L\}$, thus omitting all potential singularities at $\nu = 2\lambda_i$, $i = 1, ..., L$; the right panel displays additionally the lag-one acf 10 of the extension $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ in 9, when $\nu = \nu_{i_0} = 2\lambda_{i_0}$ for $i_0 = 8, 9, 10$, where the three additional (vertical black) spectral lines, corresponding to $\mathbf{v}_8, \mathbf{v}_9, \mathbf{v}_{10}$, show the range of acf-values as a function of $\tilde{N}_{i_0} \in \mathbb{R}$: lower and upper bounds of each spectral line correspond to $\rho_{i_0}(0) = \rho_{\nu_{i_0}} = \frac{M_{i_0 1}}{M_{i_0 2}}$, when $\tilde{N}_{i_0} = 0$ in 10, and $\rho_{i_0}(\pm\infty) = \lambda_{i_0}$, when $\tilde{N}_{i_0} = \pm\infty$. The green horizontal lines in both graphs correspond to two different arbitrary holding-times $\rho_1 = 0.6$ and $\rho_1 = 0.365$: the intersections of the latter with the acfs, marked by colored vertical lines in each panel, indicate potential solutions of the SSA-problem for the thusly specified holding-time constraint. The corresponding criterion values are reported at the bottom of the colored vertical lines: the SSA-solution is determined by the intersection which leads to the highest criterion value (rightmost in this example).

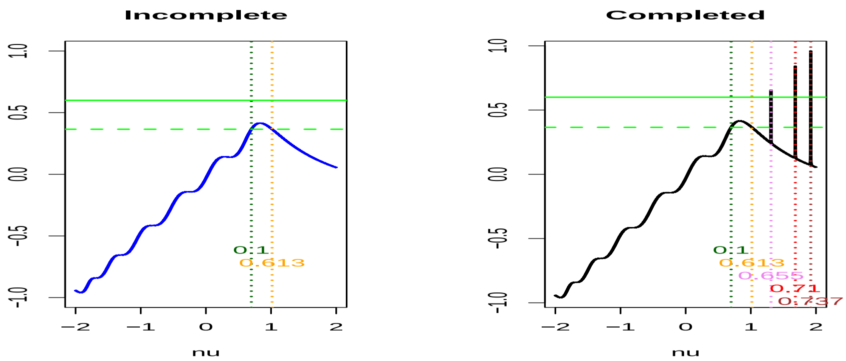

Figure 7: Lag-one autocorrelation as a function of ν. Original (incomplete) solutions (left panel) vs. completed solutions (right-panel). Intersections of the acf with the two green lines are potential solutions of the SSA-problem for the corresponding holding-times: criterion values are reported for each intersection (bottom right).

The right panel in the figure illustrates that the completion with the extensions $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ at the singular points $\nu = \nu_{i_0} = 2\lambda_{i_0}$ for $i_0 = 8, 9, 10$ can accommodate for a wider range of holding-time constraints, such that $|\rho_1| < \rho_{max}(L) = \lambda_{10} = 0.959$; in contrast, $\mathbf{b}(\nu)$ in the left panel is limited to $-0.959 = \lambda_1 < \rho_1 < \lambda_7 = 0.415$ so that there does not exist a solution for $\rho_1 = 0.6$ (no intersection with upper green line in left panel). Moreover, for a given holding-time constraint, the additional stationary points corresponding to intersections at the spectral lines of the (completed) acf might lead to improved performances, as shown in the right panel, where the maximal criterion value

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

is attained at the right-most spectral line, for $i_0 = 10$, and where $\tilde{N}_{10} = 0.077$ has been obtained from 11, with the correct signs of D and $\tilde{N}_{10}$ in place.

7.6 Proofs of Theorem (1) and Corollaries (2) and (4)

Proof of Theorem (1): **Proof**: Consider the Lagrangian $\mathcal{L} := \boldsymbol{\gamma}_\delta'\mathbf{b} - \tilde{\lambda}_1(\mathbf{b}'\mathbf{b} - 1) - \tilde{\lambda}_2(\mathbf{b}'\mathbf{M}\mathbf{b} - \rho_1)$, where we assume $l = 1$ in Criterion (1). By assumption $L \geq 3$ so that $\mathbf{b}$ is defined on a $L - 2 \geq 1$ dimensional intersection of unit-sphere and holding-time hyperbola that is free of boundary points (see the appendix for details). Therefore, the solution $\mathbf{b}$ of the SSA-problem must be a solution to the stationary Lagrangian equation

$$\boldsymbol{\gamma}_\delta = \tilde{\lambda}_1 2\mathbf{b} + \tilde{\lambda}_2(\mathbf{M} + \mathbf{M}')\mathbf{b} = \tilde{\lambda}_1 2\mathbf{b} + \tilde{\lambda}_2 2\mathbf{M}\mathbf{b}. \tag{38}$$

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

$$D\boldsymbol{\gamma}_\delta \quad = \quad \boldsymbol{\nu}\mathbf{b}, \tag{39}$$

where $\boldsymbol{\nu} := (2\mathbf{M} - \nu\mathbf{I})$, $D = 1/\tilde{\lambda}_2 \neq 0$[7] and $\nu = -2\frac{\tilde{\lambda}_1}{\tilde{\lambda}_2}$. Furthermore, Equation (39) can be written in terms of a difference equation

$$\begin{aligned} b_{k+1} - \nu b_k + b_{k-1} &= D\gamma_{k+\delta}\ ,\ 1 \leq k \leq L-2 \quad (40)\\ b_1 - \nu b_0 &= D\gamma_\delta\ ,\ k = 0\\ -\nu b_{L-1} + b_{L-2} &= D\gamma_{L-1+\delta}\ ,\ k = L-1 \end{aligned}$$

for $k = 0, ..., L-1$ so that $b_{-1} = b_L = 0$ are implicitly assumed. The eigenvalues of $\boldsymbol{\nu}$ are $2\lambda_i - \nu$ with corresponding eigenvectors $\mathbf{v}_i$. If $\mathbf{b}(\nu)$ is the solution to the SSA-problem, then $\nu/2$ cannot be an eigenvalue of $\mathbf{M}$ since otherwise $\boldsymbol{\nu}$ in 39 would map one of the eigenvectors to zero which would contradict the last regularity assumption (since $D \neq 0$). Therefore, $\nu \in \mathbb{R} \setminus \{2\lambda_i | i = 1, ..., L\}$, $\boldsymbol{\nu}^{-1}$ exists and $\boldsymbol{\nu}^{-1} = \mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'$, where the diagonal matrix $\mathbf{D}_\nu^{-1}$ has entries $\frac{1}{2\lambda_i - \nu}$. Solving for $\mathbf{b}$ in Equation (39) gives

$$\begin{aligned} \mathbf{b} &= D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta \quad (41)\\ &= D\mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'\mathbf{V}\mathbf{w}\\ &= D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i. \quad (42) \end{aligned}$$

For a proof of Assertion (2) we consider

$$\begin{aligned} \rho(\nu) &= \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\left(D\sum_{i=1}^L \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)'\mathbf{M}\left(D\sum_{i=1}^L \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)}{\left(D\sum_{i=1}^L \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)'\left(D\sum_{i=1}^L \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)}\\ &= \frac{\sum_{i=1}^L \dfrac{\lambda_i w_i^2}{(2\lambda_i - \nu)^2}}{\sum_{i=1}^L \dfrac{w_i^2}{(2\lambda_i - \nu)^2}} \quad (43) \end{aligned}$$

We infer that $\lim_{\nu \to 2\lambda_i} \rho(\nu) = \lambda_i$, $i = 1, ..., L$. Since $\lambda_L = -\rho_{max}(L)$ and $\lambda_1 = \rho_{max}(L)$, we conclude that the limits $\pm\rho_{max}(L)$ can be reached by $\rho(\nu)$. Continuity of $\rho(\nu)$ and the intermediate-value theorem then imply that any $\rho_1 \in]-\rho_{max}(L), \rho_{max}(L)[$ is admissible for the holding-time constraint.

We now proceed to a proof of Assertion (3):

$$\begin{aligned} \frac{d\rho\Big(y(\nu), y(\nu), 1\Big)}{d\nu} &= \frac{d}{d\nu}\left(\frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}\right) = \frac{d}{d\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\ \prime}\mathbf{M}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\ \prime}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}\right) = \frac{d}{d\nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}\right)\\ &= \frac{2\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta\mathbf{b}'\mathbf{b}/D - (2\mathbf{b}'\mathbf{M}\mathbf{b}/D)\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta}{((\mathbf{b}'\mathbf{b})^2/D^2)}\\ &= \frac{2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b}/D^2 - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}/D^2}{\mathbf{b}'\mathbf{b}/D^2}\\ &= 2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b} - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b} \quad (44) \end{aligned}$$

where $\boldsymbol{\nu}^{-k} := (\boldsymbol{\nu}^{-1})^k$, $\boldsymbol{\nu}^{-1\ \prime} = \boldsymbol{\nu}^{-1}$ (symmetry); commutativity of the matrix multiplications (used in deriving the third and next-to-last equations) follows from the fact that the matrices are symmetric and simultaneously diagonalizable (same eigenvectors); also we relied on generic matrix differentiation rules in the third equation[8]; finally we relied on $\mathbf{b}'\mathbf{b} = 1$ in the last equation. We

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

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

can now insert

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

which is a reformulation of $(2\mathbf{M} - \nu\mathbf{I})\boldsymbol{\nu}^{-1} = \mathbf{I}$ into the first summand in 44 to obtain

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

Plugging this expression into 44 and isolating $\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}$ gives

$$\begin{aligned}\frac{d\rho\Big(y(\nu), y(\nu), 1\Big)}{d\nu} &= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\left(2\mathbf{b}'\mathbf{M}\mathbf{b} - \nu\mathbf{b}'\mathbf{b}\right) + (\mathbf{b}'\mathbf{b})^2 \\ &= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\left(2\mathbf{M} - \nu\mathbf{I}\right)\mathbf{b} + (\mathbf{b}'\mathbf{b})^2 \\ &= -\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}\mathbf{b} + (\mathbf{b}'\mathbf{b})^2 \\ &= -\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta + (\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta)^2 \\ &= -\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-3}\mathbf{V}'\boldsymbol{\gamma}_\delta\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-1}\mathbf{V}'\boldsymbol{\gamma}_\delta + (\boldsymbol{\gamma}_\delta'\mathbf{V}\mathbf{D}^{-2}\mathbf{V}'\boldsymbol{\gamma}_\delta)^2 \\ &= -\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-3}\tilde{\boldsymbol{\gamma}}_{+\delta}\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-1}\tilde{\boldsymbol{\gamma}}_{+\delta} + (\tilde{\boldsymbol{\gamma}}_{+\delta}'\mathbf{D}^{-2}\tilde{\boldsymbol{\gamma}}_{+\delta})^2,\end{aligned} \tag{45}$$

where $\boldsymbol{\nu}^{-k} = \mathbf{V}\mathbf{D}^{-k}\mathbf{V}'$ and $\mathbf{D}^{-k}$, $k = 1, 2, 3$, is diagonal with eigenvalues $\lambda_{i\nu}^{-k} := (2\lambda_i - \nu)^{-k}$: the eigenvalues are (strictly) positive, if $\nu < -2\rho_{max}(L)$; if $\nu > 2\rho_{max}(L)$ then the eigenvalues are (strictly) negative, if k is an odd number, or (strictly) positive, if k is an even number. Finally, $\tilde{\boldsymbol{\gamma}}_{+\delta} = \mathbf{V}'\boldsymbol{\gamma}_{+\delta} = (w_1, ..., w_L)'$. We then conclude

$$\begin{aligned}\frac{d\rho\Big(y(\nu), y(\nu), 1\Big)}{d\nu} &= -\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-3}\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-1} + \left(\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-2}\right)^2 \\ &= -\sum_{i>k} w_i^2 w_k^2\Big(\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-3} + \lambda_{i\nu}^{-3}\lambda_{k\nu}^{-1} - 2\lambda_{i\nu}^{-2}\lambda_{k\nu}^{-2}\Big),\end{aligned} \tag{46}$$

where the terms in w_j^4 cancel. Consider now

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

where the strict inequality holds because $\lambda_{i\nu}^{-1} = (2\lambda_i - \nu)^{-1}$ are of the same sign, pairwise different and non-vanishing if $|\nu| > 2\rho_{max}(L)$. Since $w_i \neq 0$ (last regularity assumption: completeness) we deduce $w_i^2 w_k^2 \neq 0$ in 46. Therefore, the latter expression is strictly negative and we conclude that $\rho\Big(y(\nu), y(\nu), 1\Big)$ must be a strictly monotonic function of ν for $\nu \in \{x|x > 2\rho_{max}(L)\}$ or for $\nu \in \{x|x < -2\rho_{max}(L)\}$. From $\lim_{|\nu|\to\infty}\rho(\nu) = \frac{\sum_{i=1}^L \lambda_i w_i^2}{\sum_{i=1}^L w_i^2} = \rho_{MSE}$ and $\frac{d\rho(\nu)}{d\nu} < 0$ we then infer $\max_{\nu<-2\rho_{max}(L)}\rho(\nu) = \min_{\nu>2\rho_{max}(L)}\rho(\nu) = \rho_{MSE}$.
For a proof of the last Assertion (4) we first consider

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

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

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

The last equality is obtained by recognizing that the expression in curly brackets is the same as in Equation (45). If $\nu > 2\rho_{max}(L)$ the same proof applies, but with changed sign, $\mathrm{sign}(D) = -1$, and accordingly modified strict inequality. $\square$

Remarks: The length constraint linearizes the optimization problem in the sense that (non-linear) correlation and autocorrelation of Criterion (2) can be replaced by (linear) covariance and (quadratic) autocovariance in Criterion (1). Since $D = 1/\tilde{\lambda}_2$ we infer that the holding time constraint interferes with the length constraint; however, from Criterion (2) we deduce that the length constraint interferes neither with the objective function nor with the holding time constraint. Therefore, the scaling s such that $s^2\mathbf{b}'\mathbf{b} = l$ is a nuisance parameter which can be set afterwards, once a solution of arbitrary length has been determined: this facility motivates the choice of Criterion (1), or Criterion (2), over a classic MSE approach in Wildi (2024). If required, the optimal (MSE-) scaling is obtained as $s_{MSE} := \mathbf{b}'\boldsymbol{\gamma}_\delta/\mathbf{b}'\mathbf{b}$.

Proof of Corollary (2): The first assertion follows directly from the Lagrangian Equation (39). Under the case posited in the second assertion, $\boldsymbol{\nu}_{i_0}$ is not of full rank and $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ as defined by Equation (9) is a solution to the Lagrangian equation $D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}_{i_0}\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ for arbitrary $\tilde{N}_{i_0}$. Moreover,

$$\rho_{i_0}(\tilde{N}_{i_0}) := \frac{\mathbf{b}_{i_0}(\tilde{N}_{i_0})'\mathbf{M}\mathbf{b}_{i_0}(\tilde{N}_{i_0})}{\mathbf{b}_{i_0}'(\tilde{N}_{i_0})\mathbf{b}_{i_0}(\tilde{N}_{i_0})} = \frac{\sum_{i\neq i_0}\lambda_i w_i^2\frac{1}{(2\lambda_i-\nu)^2} + \tilde{N}_{i_0}^2\lambda_{i_0}}{\sum_{i\neq i_0} w_i^2\frac{1}{(2\lambda_i-\nu)^2} + \tilde{N}_{i_0}^2} = \frac{M_{i_01} + \tilde{N}_{i_0}^2\lambda_{i_0}}{M_{i_02} + \tilde{N}_{i_0}^2}.$$

Solving for the holding-time constraint $\rho_{i_0}(\tilde{N}_{i_0}) = \rho_1$ then leads to $N_{i_0} := \tilde{N}_{i_0}^2 = \frac{\rho_1 M_{i_02} - M_{i_01}}{\lambda_{i_0} - \rho_1}$. We infer that N_{i_0} is positive if $0 < \rho(\nu_{i_0}) = \frac{M_{i_01}}{M_{i_02}} < \rho_1 < \lambda_{i_0}$ or $0 > \rho(\nu_{i_0}) = \frac{M_{i_01}}{M_{i_02}} > \rho_1 > \lambda_{i_0}$, so that $\tilde{N}_{i_0} = \pm\sqrt{N_{i_0}} \in \mathbb{R}$, as claimed. Finally, the correct sign combination of the pair $D, \tilde{N}_{i_0}$ is determined by the maximal criterion value. For a proof of the third and last assertion we first assume that $\boldsymbol{\gamma}_\delta$ is not band-limited so that $w_1 \neq 0$ and $w_L \neq 0$. Then, $\lim_{\nu\to 2\lambda_1}\rho(\nu) = \lambda_1 = \rho_{max}(L)$ and $\lim_{\nu\to 2\lambda_L}\rho(\nu) = \lambda_L = -\rho_{max}(L)$, see the proof of Theorem (1). By continuity of $\rho(\nu)$ and by virtue of the intermediate-value theorem, any ρ_1 such that $|\rho_1| < \rho_{max}(L)$ is admissible for the holding-time constraint. Otherwise, if $w_1 = 0$ then $\mathbf{b}_1(\tilde{N}_1)$, where $i_0 = 1$ in Equation (9), can 'fill the gap' and reach out the lower boundary $-\rho_{max}(L)$ as $\tilde{N}_1 \to \infty$. A similar reasoning would apply in the case $w_L = 0$. $\square$

Proof of Corollary (4): Let $\gamma_{k+\delta} = \lambda^{k+\delta}$. Then

$$b_k' := D\frac{\lambda^\delta}{\lambda^2 - \nu\lambda + 1}\lambda^{k+1} \propto \lambda^{k+\delta} \tag{47}$$

is a solution of

$$b'_{k+1} - \nu b'_k + b'_{k_1} \quad = \quad D\gamma_{k+\delta}\ ,\ 0 \le k \le L-1$$

with boundaries $b_{-1}, b_L \neq 0$. The expression is well-defined because $\lambda^2 - \nu\lambda + 1 \neq 0$ when $\lambda \neq \lambda_{1\rho_1}$, by assumption. According to Proposition (4) vanishing boundaries $b_{-1} = b_L = 0$ can be imposed by combining b'_k in 47 with a suitably scaled solution of the homogeneous difference-equation:

$$b_k = b_k(\lambda_{1\rho_1}) \propto \lambda^{k+\delta} + C_1\lambda_{1\rho_1}^k + C_2\lambda_{1\rho_1}^{L-k} \approx \lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k \tag{48}$$

where we neglected the backward-solution ($b_L \approx 0$ by the last assumption) and where the weight $C_1 = -\lambda_{1\rho_1}\lambda^{-1+\delta}$ ensures compliance with the remaining constraint $b_{-1} = 0$. Corollary (3) states that the unknown stable root $\lambda_{1\rho_1}$ is determined uniquely by requiring

$$\frac{\sum_{k=1}^{L-1} b_k(\lambda_{1\rho_1}) b_{k-1}(\lambda_{1\rho_1})}{\sum_{k=0}^{L-1} b_k(\lambda_{1\rho_1})^2} = \rho_1 \tag{49}$$

We can now insert 48 into this equation and solve for $\lambda_{1\rho_1}$. Specifically, the nominator becomes

$$\sum_{k=1}^{L-1} b_k b_{k-1} \quad = \quad \sum_{k=1}^{L-1} \left(\lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k\right)\left(\lambda^{k-1+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^{k-1}\right) \tag{50}$$

The first cross-product of terms in parantheses is

$$\lambda^{-1}\sum_{k=1}^{L-1}\lambda^{2(k+\delta)} = \lambda^{1+2\delta}\sum_{k=0}^{L-2}\lambda^{2k} = \lambda^{1+2\delta}\frac{1-\lambda^{2(L-1)}}{1-\lambda^2} \approx \frac{\lambda^{1+2\delta}}{1-\lambda^2} \tag{51}$$

The second cross-product of terms in parentheses is

$$-\lambda_{1\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k+\delta}\lambda_{1\rho_1}^{k-1} = -\lambda_{1\rho_1}\lambda^{2\delta}\sum_{k=0}^{L-2}(\lambda\lambda_{1\rho_1})^k = -\lambda_{1\rho_1}\lambda^{2\delta}\frac{1-(\lambda\lambda_{1\rho_1})^{L-1}}{1-\lambda\lambda_{1\rho_1}} \approx \frac{-\lambda_{1\rho_1}\lambda^{2\delta}}{1-\lambda\lambda_{1\rho_1}} \tag{52}$$

The third cross-product of terms in parentheses is

$$-\lambda_{1\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k-1+\delta}\lambda_{1\rho_1}^k \approx \frac{-\lambda_{1\rho_1}^2\lambda^{2\delta-1}}{1-\lambda\lambda_{1\rho_1}} \tag{53}$$

The last cross-product of terms in parantheses is

$$\lambda_{1\rho_1}^2\lambda^{-2+2\delta}\sum_{k=1}^{L-1}\lambda_{1\rho_1}^{2k-1} = \lambda_{1\rho_1}^3\lambda^{-2+2\delta}\sum_{k=0}^{L-2}\lambda_{1\rho_1}^{2k} = \lambda_{1\rho_1}^3\lambda^{-2+2\delta}\frac{1-\lambda_{1\rho_1}^{2(L-1)}}{1-\lambda_{1\rho_1}^2} \approx \frac{\lambda_{1\rho_1}^3\lambda^{-2+2\delta}}{1-\lambda_{1\rho_1}^2} \tag{54}$$

Note that the last assumption of the corollary is critical for the validation of the above approximations. Further, the common denominator of 51, 52, 53 and 54 is

$$(1-\lambda^2)(1-\lambda\lambda_{1\rho_1})(1-\lambda_{1\rho_1}^2) \tag{55}$$

Summing all terms in 51, 52, 53 and 54 under the common denominator 55 leads to a third-order polynomial

$$f_1(\lambda_{1\rho_1}) := a_3\lambda_{1\rho_1}^3 + a_2\lambda_{1\rho_1}^2 + a_1\lambda_{1\rho_1} + a_0$$

in $\lambda_{1\rho_1}$ with coefficients

$$\begin{aligned}
a_3 &= (1-\lambda^2)\lambda^{2\delta-2}(\lambda^2+1) + \lambda^{2+\delta} = \lambda^{2\delta-2} - \lambda^{2\delta+2} + \lambda^{2+\delta} \\
a_2 &= -(\lambda^{1+\delta} + \lambda^{2\delta-1}(1-\lambda^2)) \\
a_1 &= -(\lambda^{2+\delta} + \lambda^{2\delta}(1-\lambda^2)) \\
a_0 &= \lambda^{1+\delta}
\end{aligned}$$

Note that the coefficient of order four vanishes due to the mutual cancellation of cross-terms. The same proceeding can now be applied to the denominator $\sum_{k=0}^{L-1} b_k^2$ in 49:

$$\sum_{k=0}^{L-1} b_k^2 = \sum_{k=0}^{L-1} \left(\lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k\right)^2 \tag{56}$$

with cross-products of terms in parentheses

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

This will again lead to the sum of four terms whose common denominator is again 55 and whose nominator is a polynomial

$$f_2(\lambda_{1\rho_1}) := b_3\lambda_{1\rho_1}^3 + b_2\lambda_{1\rho_1}^2 + b_1\lambda_{1\rho_1} + b_0$$

with polynomial coefficients

$$\begin{aligned}
b_3 &= \lambda^{2\delta-1} \\
b_2 &= -2\lambda^{2\delta} + \lambda^{2\delta-2} \\
b_1 &= \lambda^{2\delta+1} - 2\lambda^{2\delta-1} \\
b_0 &= \lambda^{2\delta}
\end{aligned}$$

Note that fourth-order terms do not appear in this case. After cancellation of their common denominator 55, equation 49 can then be re-written as

$$\frac{f_1(\lambda_{1\rho_1})}{f_2(\lambda_{1\rho_1})} = \rho_1$$

or

$$f_3(\lambda_{1\rho_1}, \rho_1) = 0 \tag{57}$$

where $f_3(\lambda_{1\rho_1}, \rho_1) := f_1(\lambda_{1\rho_1}) - \rho_1 f_2(\lambda_{1\rho_1})$ is the asserted cubic polynomial in $\lambda_{1\rho_1}$ with coefficients

$$c_3 = a_3 - \rho_1 b_3 \ , \ c_2 = a_2 - \rho_1 b_2 \ , \ c_1 = a_1 - \rho_1 b_1 \ , \ c_2 = a_0 - \rho_1 b_0$$

The remainder of the proof then follows from a closed-form expression for the root of a cubic polynomial[9], see e.g. Cardano's formula.

Remarks

In contrast to the *frequency-domain* decomposition 5 of the lag-one autocorrelation, the above *time-domain* decomposition allows for notable simplifications resulting in a closed-form solution for $\lambda_{1\rho_1}$ under the posited assumptions. The above results can be generalized to the non-degenerate singular case, when $\mathbf{b} \not\propto \boldsymbol{\gamma}_\delta$ but $\lambda = \lambda_{1\rho_1}$. Also, extensions to an AR(2)-target z_t are available in closed-form but omitted here due to more convoluted algebraic expressions[10]. Note also that if $|\nu| < 2$ then $\lambda_{1\rho_0}$ is a unit-root so that (some of) the geometric series in the above proof do not converge anymore and therefore high-order terms (power L) cannot be neglected anymore. The resulting inflated polynomial order is an indication of multiple solutions for ν given ρ_1 in this case.

[9] Under particular circumstances the leading polynomial coefficient can vanish, $c_3 = 0$, so that the solution for $\lambda_{1\rho_1}$ simplifies to finding the root of a quadratic polynomial.

[10] Additional complexity arises because the weights of the MA-decomposition of a stationary AR(p)-process are not an eigenfunction of the AR(2)-operator 40, unlike the discussed AR(1)-case, see equation 47.

References

[1] Anderson O.D. (1975) Moving Average Processes. *Journal of the Royal Statistical Society. Series D (The Statistician).* **Vol. 24, No. 4**, 283-297

[2] Barnett J.T. (1996) Zero-crossing rates of some non-Gaussian processes with application to detection and estimation. *Thesis report Ph.D.96-10, University of Maryland.*

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

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

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

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

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

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

[9] 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.

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

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

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