# Multivariate Real-Time Signal Extraction

Marc Wildi and Tucker McElroy

February 2, 2023

# Contents

# Chapter 1

# Introduction

## 1.1 Overview

### 1.1.1 Signals and Extraction

In the applications of time series analysis to macroeconomics, finance, and quality control it is essential to extract useful information about trends, turning points, and anomalies in real time. The practitioner does not have the luxury of sifting past data for structural breaks, indicators of regime change, or changes to volatility. Informative elections are contingent upon understanding the dynamics of various time series at time present. Because long-term movements, as well as aberrations, are defined in terms of the long-run behavior of a time series over past, present, and future, any analysis of the present state necessarily involves a degree of forecasting. This broad topic is referred to as real-time signal extraction.

A signal is any component of a time series that is deemed useful for a particular application. If long-term movements are of interest, the signal is a trend. If short-term fluctuations about a longer-term mean are of interest, the signal is a cycle. If shocks, due to rare terrorist events or natural disasters, are of interest, the signal consists of the extreme values. If regular patterns of an annual period, linked to cultural or meteorological patterns, are of interest, the signal is a seasonal component.

However, these signals are not directly observable at time present, because in each case their definition involves all the past and future values of a time series – but the future is unknown, and only part of the past is available to us. The statistical processes by which a signal is estimated from available data is referred to as extraction, and the residual from the signal extraction is referred to as the noise. Whereas signals can be estimated from historical, or past, sections of a time series, when effort is focused upon time present we refer to the analysis as real-time signal extraction.

Real-time signal extraction is considerably more challenging, and useful, than historical signal extraction. The difficulty lies in the uncertainty about the future, which is transmitted unto the signal extraction estimates themselves. One way to conceive of this difficulty is through the warring principles of timeliness and accuracy: should we procrastinate in providing our analysis of the present, we can increase the accuracy of signal extraction, but our answers become less

relevant, even as the present time rapidly drifts into the past. Conversely, extremely timely extractions suffer from greater future uncertainty, and are likely to exhibit inaccuracy.

There is a considerable body of literature addressing signal extraction, but this book focuses upon a particular methodology called Direct Filter Analysis (DFA). As the original development of DFA was univariate, the methodology's power was limited to the information content within a single time series. But because batches of time series can be closely linked, exhibiting correlated trends, common dynamics, or even predictive relationships, it is natural to expect that a multivariate extension of DFA to vector time series will more greatly facilitate informed decision making. The topic of this book is Multivariate Direct Filter Analysis (MDFA).

### 1.1.2 The Classic Model-Based Paradigm

Many signals can be formulated as weighted linear combinations of a time series, in which case the real-time signal extraction problem can be approached as a Linear Prediction Problem (LPP). In order to pose an LPP, a solution criterion is needed, and Mean Squared Error (MSE) is often used: one seeks a real-time signal extraction that has minimal MSE discrepancy with the actual target signal. Although an LPP can then be solved, the solution depends on knowing something about the dynamics in the time series process. The most venerable approach to understanding these dynamics is to posit a time series model, and fit this model to the observed data. This approach, which goes back to the work of Yule in the 1930s, is called the classic paradigm, being based upon a Model-Based Analysis (MBA).

An attractive feature of MBA is that analytical formulas for the LPP solutions can often be obtained, thereby facilitating computation. The philosophy underpinning the classic paradigm is that a Data Generation Process (DGP) exists – as a theoretical, or Platonic construct – to which the observed data closely adheres. Formally, the DGP is some stochastic process defined upon a probability space, and the observed data is a realization, or sample path, of the DGP. Statistical inference is involved with the science of identifying a model class for the DGP, narrowing down the class to a particular model (by eliminating contenders), and fitting that model via fixing values of the parameters. Successive applications of model diagnostics allow for refinements, and a process by which we can verify the validity of a postulated model. Of course, all of this is done on the basis of the single realization of the DGP.

While recognizing that any such model need not be correct, i.e., exactly match the DGP itself, such models can yet be useful to the extent to which they reflect important features in the data. Yet it is difficult to keep a model simple – which is necessay to its utility – and at the same time be sufficiently versatile to explain all the data's features. Moreover, the appellation of importance is subjective: a feature deemed important to one user may be irrelevant to another. This begs the question of customization: each user, with a distinct set of criteria and desired applications, could potentially stress the importance of a subset of features at the cost of de-emphasizing others. The classic paradigm ignores, or at least passes over, the issue of customization, and proposes a single all-purpose concept of utility: the minimization of one-step ahead forecast error MSE.

Another term for this classic conception of model utility is the Wold decomposition, which breaks a wide class of stochastic processes down in terms of a component that is completely

predictable from its own infinite past, and a second component fully describable in terms of one-step ahead forecast errors. Classical models can then be viewed as attempts to approximate the linear machinery in the Wold decomposition. However, were attention to focus upon an alternative utility, e.g., 10-step ahead forecasting, a different class of models would be suggested, with different apparatus for model selection, fitting, and evaluation.

However, customizing the modeling apparatus to allow for specific applications offers only a partial solution, because model mis-specification is the larger challenge. The full set of LPP solutions for a given time series is greatly constrained once a model is introduced, as only a particular subset of solutions can be obtained. If the model is badly mis-specified, the resulting LPP solution will be inadequate, even if the criteria for model selection are customized. This empirical disfunctionality motivated the genesis of DFA, which essentially provides access to a much wider pool of LPP solutions. Moreover, the basic DFA can be easily modified to allow for direct customization of real-time problems, according to whether users are concerned with timeliness, accuracy, or fidelity to the original signal (called smoothness).

### 1.1.3 The Scope of MDFA

Our critique of the classic paradigm has several facets. First, there is typically model mis-specification present. Second, the problem has typically not been structured properly, in the sense that the criteria used do not correspond to the relevant LPP, but rather to one-step ahead forecasting. Third, there is no specific customization of the model, in order to account for timeliness and accuracy. These weaknesses are actually linked together.

Model mis-specification is always present; the issue is whether it has a significant impact upon the objectives of analysis. For instance, a given model's mis-specification may have grave repercussions for certain problem structures, while being adequate for other LPPs. The given LPP of interest determines the gravity and impact of model mis-specification. Moreover, in the classic paradigm the one-step ahead forecasting LPP is solved, and it is merely hoped that timeliness and accuracy will be adequate for all users. Model parameters can be tweaked, or tuned, in order to indirectly modify timeliness and accuracy – but the relationships are indirect and often poorly understood. By building the timeliness-accuracy tradeoff directly into the DFA criterion, the optimality of an LPP solution for a customized application is assured.

These topics have been treated in Wildi and McElroy (2016) in the case of univariate time series, which discusses at length the basic DFA. This book presents the generalized treatment of the multivariate LPP in Chapter **??**. But before discussing customization in Chapter **??**, we discuss the applications of forecasting and filtering in Chapter **??**, wherein the multivariate DFA is introduced in the context of stationary processes. Then this basic treatment is extended to nonstationary processes in Chapter **??**, including a discussion of filter constraints; Chapter **??** extends the methodology to the case of co-integration. These chapters also provide applications to replicating and enhancing classical model-based approaches, as well as Hodrick-Prescott and band-pass filters. Additional topics include zero-crossings (Chapter **??**) and mixed-frequency data (Chapter **??**).

## 1.2 The Style of the Book

This book was generated using Sweave, in accordance with the philosophy of scientific replicability. Throughout the text are portions of R code that can be pasted into an R script and directly run, given that the user has certain packages already installed. This installation is described below.

### 1.2.1 Setting the Paths

Begin by clearing the workspace:

```
> #rm(list=ls())
```

The R code in various chapters of this book requires installation of the following R packages:

```
> # Load packages: time series and xts
> library(xts)  #library(tseries)
> library(dlm)  # State-space models (will be replicated by MDFA)
> library(mFilter)  # Classic filter designs (will be replicated by MDFA)
> library(numDeriv) # Numerical package
> library(tis)  # Graphical package for recession-shading
> require(xtable) # Library for tables
> library(devtools)
> # Load MDFA package from github
> devtools::install_github("wiaidp/MDFA")
> library(MDFA) # MDFA package
```

US-GDP data for the empirical examples can be retrieved either directly from Quandl (requiring a preliminary user registration) or from a local data folder, which is the default-setting:

```
> # Load fresh data from quandl: T/F
> #   Default-setting is False: the data will be loaded from local data folder
> load_from_quandl <- F
```

Paths to MDFA code, as well as to the US-GDP data, must be provided by the reader. It is assumed that the MDFA package is saved to a main folder containing subfolders labeled as DFA, MDFA, model-based, and data. The R code in the book generates pdf graphs that are saved in a separate folder, whose path is specified by *path.out*.

```
> path.main <- paste(getwd(),"/Sweave/",sep="") # Set main path
> #path.main <- "C:\\Users\\Tucker\\Documents\\MDFAbook\\"
> # Set paths to subfolders
>   # Path to Latex-folder: all pdfs generated by the R code are filed there
> path.out <- paste(path.main,"Latex/",sep="")
>   # Path to data (US-GDP)
> path.dat <- paste(path.main,"Data/",sep="")
>   # Path to code that is part of MDFA-Legacy project but not part of MDFA package
> path.pgm <- paste(path.main,"R/",sep="")
```

HERE: following sections can be removed, they are covered in the text

### 1.2.2 DFA

We here briefly review the relevant facets of DFA, thereby providing an anchor for the MDFA discussion.

#### DFT and Periodogram

The Discrete Fourier Transform (DFT) and the periodogram are defined in Sections 2.2 and 2.3 of DFA. The following periodogram function – referred to as *per* below – in the MDFA package replicates these formulae. Note that frequency $\pi$ is treated differently, depending on whether the sample size is odd or even; also, the value at frequency zero is scaled by $1/\sqrt{2}$, which is explained in later text.

```
> head(per,100)
```

```
1  function (x, plot_T)
2  {
3      len <- length(x)
4      per <- 0:(len/2)
5      DFT <- per
6      for (k in 0:(len/2)) {
7          cexp <- exp((0+1i) * (1:len) * 2 * pi * k/len)
8          DFT[k + 1] <- sum(cexp * x * sqrt(1/(2 * pi * len)))
9      }
10     if (abs(as.integer(len/2) - len/2) < 0.1)
11         DFT[k + 1] <- DFT[k + 1]/sqrt(2)
12     per <- abs(DFT)^2
13     if (plot_T) {
14         par(mfrow = c(2, 1))
15         plot(per, type = "l", axes = F, xlab = "Frequency", ylab = "Periodogram",
16             main = "Periodogram")
17         axis(1, at = 1 + 0:6 * len/12, labels = c("0", "pi/6",
18             "2pi/6", "3pi/6", "4pi/6", "5pi/6", "pi"))
19         axis(2)
20         box()
21         plot(log(per), type = "l", axes = F, xlab = "Frequency",
22             ylab = "Log-periodogram", main = "Log-periodogram")
23         axis(1, at = 1 + 0:6 * len/12, labels = c("0", "pi/6",
24             "2pi/6", "3pi/6", "4pi/6", "5pi/6", "pi"))
25         axis(2)
26         box()
27     }
```

```
28     return(list(DFT = DFT, per = per))
29 }
```

This function will be generalized in the new multivariate setting.

**Basic DFA**

A simple version of the DFA based on the MSE criterion alone – as proposed in Section 4.1 of DFA – is included in the MDFA package:

```
> # This function computes MSE DFA solutions
> # L is the length of the MA filter,
> # periodogram is the frequency weighting function in the DFA
> # Gamma is the transfer function of the symmetric filter (target) and
> # Lag is the lag-parameter: Lag=0 implies real-time filtering, Lag=L/2
> #     implies symmetric filter
> # The function returns optimal coefficients as well as the transfer
> #     function of the optimized real-time filter
> head(dfa_ms,100)

1  function (L, periodogram, Lag, Gamma)
2  {
3      periodogram[1] <- periodogram[1]/2
4      K <- length(periodogram) - 1
5      X <- exp(-(0+1i) * Lag * pi * (0:(K))/(K)) * rep(1, K + 1) *
6          sqrt(periodogram)
7      X_y <- exp(-(0+1i) * Lag * pi * (0:(K))/(K)) * rep(1, K +
8          1)
9      for (l in 2:L) {
10         X <- cbind(X, (cos((l - 1 - Lag) * pi * (0:(K))/(K)) +
11             (0+1i) * sin((l - 1 - Lag) * pi * (0:(K))/(K))) *
12             sqrt(periodogram))
13         X_y <- cbind(X_y, (cos((l - 1 - Lag) * pi * (0:(K))/(K)) +
14             (0+1i) * sin((l - 1 - Lag) * pi * (0:(K))/(K))))
15     }
16     xtx <- t(Re(X)) %*% Re(X) + t(Im(X)) %*% Im(X)
17     b <- as.vector(solve(xtx) %*% (t(Re(X_y)) %*% (Gamma * periodogram)))
18     trffkt <- 1:(K + 1)
19     trffkt[1] <- sum(b)
20     for (k in 1:(K)) {
21         trffkt[k + 1] <- (b %*% exp((0+1i) * k * (0:(length(b) -
22             1)) * pi/(K)))
23     }
```

```
24     return(list(b = b, trffkt = trffkt))
25 }
```

This function is nested in the multivariate MDFA, in the sense that the latter can replicate the former perfectly when suitably parametrized; see Section **??** below.

#### Customized DFA

A more general DFA function, called *dfa_analytic*, is proposed in Section 4.3.5 of DFA. Customization and the generic Accuracy-Timeliness-Smoothness (ATS) trilemma are presented in Sections 4.3 and 5 of DFA. This function is included in the MDFA package:

```
> head(dfa_analytic)
```

```
1 function (L, lambda, periodogram, Lag, Gamma, eta, cutoff, i1,
2     i2)
3 {
4     periodogram[1] <- periodogram[1]/2
5     lambda <- abs(lambda)
6     eta <- abs(eta)
```

The additional control parameters *lambda*, *eta* allow for customization of the filter, as discussed below in Chapter **??**. The Boolean *i1* and *i2* can enforce useful filter constraints; see Chapter **??**. This function is also encompassed by the MDFA.

### 1.2.3 MDFA

The R code for MDFA is more sophisticated than that of the DFA, and is correspondingly more complex and lengthy. As for the DFA package, the MDFA code can be sourced. We here briefly review the corresponding pieces.

#### Data Matrix

All time series are collected in a data-*matrix*, say $X$, which is organized as follows:

- the first column $X[,1]$ of $X$ always corresponds to the target series: the target series $X[,1]$ is the time series to be forecasted, nowcasted or backcasted.
- Columns 2, 3, ... of $X$ are allocated to the explanatory variables (more than one in a multivariate setting). If the target series is part of the set of explanatory variables (it does not have to be), then it must be assigned a specific column – by convention always the second one – in $X$, i.e., in this case the target series is entered twice, in the first column (target) and in the second column (explanatory data).

**Example**. Suppose we study a two-dimensional signal extraction problem, whereby the target series (first column) is part of the set of explanatory variables:

```
> set.seed(1)
> len <- 100
> target <- arima.sim(list(ar=0.9),n=len)
> explanatory_2 <- target+rnorm(len)
> explanatory <- cbind(target,explanatory_2)
> x <- cbind(target,explanatory)
> dimnames(x)[[2]] <- c("target","explanatory 1","explanatory 2")
> head(x)

      target explanatory 1 explanatory 2
[1,] 1.703613      1.703613     0.3191863
[2,] 1.398197      1.398197     3.2674879
[3,] 3.659995      3.659995     4.0850957
[4,] 3.254756      3.254756     3.0161086
[5,] 3.619020      3.619020     4.6775026
[6,] 3.285120      3.285120     4.1715424
```

For a one-step ahead forecast LPP, we might consider lagging both the explanatory variables:

```
> x<-cbind(x[,1],lag(x[,2:3],-1))
> dimnames(x)[[2]]<-c("target","lagged explanatory 1","lagged explanatory 2")
> head(x)

      target lagged explanatory 1 lagged explanatory 2
[1,] 1.703613                  NA                   NA
[2,] 1.398197            1.703613            0.3191863
[3,] 3.659995            1.398197            3.2674879
[4,] 3.254756            3.659995            4.0850957
[5,] 3.619020            3.254756            3.0161086
[6,] 3.285120            3.619020            4.6775026
```

By adopting the frequency-domain methods of this book, we can generalize this construction and avoid the introduction of missing values (denoted by NA in R). □

**DFT**

In contrast to the univariate DFA, where the LPP can be expressed in terms of the periodogram, the multivariate case requires the DFT of each time series in order to account for cross-sectional dependencies. These DFTs are complex-valued quantities, and the angular portion of the cross-spectrum provides information about the relative phase-shift of each explanatory time series. In the univariate case the relative phase-shift is irrelevant, because the target series and the explanatory series are identical. The scope of the method is extended in order to cover the mixed-frequency case, which is discussed in Chapter **??**. Another facet, is that we allow for the possibility of integrated processes; see Chapter **??**. In order to illustrate some of the new features we briefly look at the main DFT function called *spec_comp*:

```
> spec_comp

function (insamp, x, d)
{
    if (d == 1) {
        weight_func <- periodogram_bp(diff(x[1:insamp, 1]), 1,
            insamp - 1)$fourtrans
        if (length(weight_func) > 1) {
            for (j in 2:ncol(x)) {
                per <- periodogram_bp(diff(x[1:insamp, j]), 1,
                  insamp - 1)$fourtrans
                weight_func <- cbind(weight_func, per)
            }
        }
    }
    else {
        weight_func <- periodogram_bp(x[1:insamp, 1], 0, insamp)$fourtrans
        if (length(weight_func) > 1) {
            for (j in 2:ncol(x)) {
                per <- periodogram_bp(x[1:insamp, j], 0, insamp)$fourtrans
                weight_func <- cbind(weight_func, per)
            }
        }
    }
    colnames(weight_func) <- colnames(x)
    return(list(weight_func = weight_func))
}
<bytecode: 0x000001ddbaf5fae8>
<environment: namespace:MDFA>
```

The inner loop tracks the columns of the data matrix $X$ and the DFTs are stored in a matrix called *weight_func*, which is returned by the function. The matrix *weight_func* collects all DFTs; the target series is always in the first column, whereas the DFTs of the explanatory series are in columns 2, 3, ... The function *periodogram_bp*, called in the above loop, is slightly more general than the DFA function *per* proposed in the previous section. In particular, it can handle various integration orders as well as seasonal peculiarities.

### 1.2.4 Using MDFA

**A Versatile User Interface**

MDFA is a generic forecast and signal extraction paradigm. Besides its capacity to replicate classical time series approaches, MDFA possesses unique features such as customization and regularization (Chapter **??**); it can treat data revisions (Chapter **??**), mixed-frequency problems

(Chapter **??**), and non-stationarity (Chapters **??** and **??**. Accordingly, the user interface is more sophisticated than the precediing DFA package. Consider the head of the main estimation routine:

```
> head(mdfa_analytic)
```

```
1 function (L, lambda, weight_func, Lag, Gamma, eta, cutoff, i1,
2     i2, weight_constraint, lambda_cross, lambda_decay, lambda_smooth,
3     lin_eta, shift_constraint, grand_mean, b0_H0, c_eta, weight_structure,
4     white_noise, synchronicity, lag_mat, troikaner)
5 {
6     lambda <- abs(lambda)
```

Arguments such as *weight_func* (discussed above), the filter length ($L$), and the target specification *Gamma* are straightforward. But there are numerous additional control parameters: the relevance and the modus operandi of these will be discussed in this book.

**Default Settings**

For convenience, we store a so-called default setting of the parameters in a file called *control_default*. First we define the data (initialize the DFT matrix) and specify the filter length:

```
> weight_func <- matrix(rep(1:6,2),ncol=2)
> L <- 2
```

Given these two entries (DFT and filter length), the default-settings are as follows:

```
> d<-0
> lin_eta<-F
> lambda<-0
> Lag<-0
> eta<-0
> i1<-F
> i2<-F
> weight_constraint<-rep(1/(ncol(weight_func)-1),ncol(weight_func)-1)
> lambda_cross<-lambda_smooth<-0
> lambda_decay<-c(0,0)
> lin_expweight<-F
> shift_constraint<-rep(0,ncol(weight_func)-1)
> grand_mean<-F
> b0_H0<-NULL
> c_eta<-F
> weights_only<-F
> weight_structure<-c(0,0)
> white_noise<-F
> synchronicity<-F
```

```
> cutoff<-pi
> lag_mat<-matrix(rep(0:(L-1),ncol(weight_func)),nrow=L)
> troikaner<-F
```

This particular configuration will be used extensively in Chapter **??**; it corresponds to the basic MSE criterion (i.e., no customization) without regularization, without design constraints, and without any *a priori* knowledge. Also, this configuration presumes a common identical sampling frequency (i.e., no mixed frequency data) and the absence of data revisions. The default settings can be obtained by sourcing the corresponding R file:

```
> source(file=paste(path.pgm,"control_default.r",sep=""))
```

For later use we source a convenient plotting function:

```
> source(file=paste(path.pgm,"mplot_func.r",sep=""))
```

**Selected Calls: Classic MSE, Customization and Regularization**

Selected calls of the classic MSE criterion – as well as calls utilizing the customization or regularization features – are available through dedicated functions in the MDFA package:

```
> head(MDFA_mse)

1 function (L, weight_func, Lag, Gamma)
2 {
3     cutoff <- pi
4     lin_eta <- F
5     lambda <- 0
6     eta <- 0

> head(MDFA_mse_constraint)

1 function (L, weight_func, Lag, Gamma, i1, i2, weight_constraint,
2     shift_constraint)
3 {
4     cutoff <- pi
5     lin_eta <- F
6     lambda <- 0

> head(MDFA_cust)

1 function (L, weight_func, Lag, Gamma, cutoff, lambda, eta)
2 {
3     lin_eta <- F
4     weight_constraint <- rep(1/(ncol(weight_func) - 1), ncol(weight_func) -
5         1)
6     lambda_cross <- lambda_smooth <- 0
```

```
> head(MDFA_cust_constraint)

1 function (L, weight_func, Lag, Gamma, cutoff, lambda, eta, i1,
2     i2, weight_constraint, shift_constraint)
3 {
4     lin_eta <- F
5     lambda_cross <- lambda_smooth <- 0
6     lambda_decay <- c(0, 0)

> head(MDFA_reg)

1 function (L, weight_func, Lag, Gamma, cutoff, lambda, eta, lambda_cross,
2     lambda_decay, lambda_smooth, troikaner = F, b0_H0 = NULL)
3 {
4     lin_eta <- F
5     weight_constraint <- rep(1/(ncol(weight_func) - 1), ncol(weight_func) -
6         1)

> head(MDFA_reg_constraint)

1 function (L, weight_func, Lag, Gamma, cutoff, lambda, eta, lambda_cross,
2     lambda_decay, lambda_smooth, i1, i2, weight_constraint, shift_constraint,
3     troikaner = F, b0_H0 = NULL)
4 {
5     lin_eta <- F
6     lin_expweight <- F
```

The heads of the corresponding functions differ in the number of additional arguments available when going from specific (MSE) to generic (reg). The following chapters of the book provide an understanding of the use of these functions.

# Chapter 2

# Mean-Square Error, Zero-Crossings, and Sign Accuracy

## 2.1 Introduction

Time series forecasting aims at a coherent description of the main systematic dynamics of a phenomenon in view of synthesizing information about future events. Typically, the forecast process is structured by a formal optimality concept, whereby a particular forecast-error measure, such as e.g. the mean-square error (MSE), is minimized. We here argue that alternative characteristics of a predictor might draw attention such as the smoothing capability, i.e. the extent by which undesirable 'noisy' components of a time series are suppressed, or timeliness, as measured by leading properties of a predictor, or sign accuracy and zero-crossings, as measured by the ability to predict the correct sign of the target. For that purpose, we here propose a generic forecast approach by merging 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 in a fixed time interval. Zero-crossings (of the growth-rate) of a time series are influential in the decision-making, for e.g. economic actors, by marking transitions between up- and down-turns, expansions and recessions, and our forecast approach contributes to such a design of the predictor. McElroy and Wildi (2019) propose an alternative methodological framework for addressing specific facets of the forecast problem but their approach does not account for zero-crossings explicitly which may be viewed as a shortcoming in some applications. Wildi (2023) illustrates application-side aspects of the novel approach, in a (real-time) business-cycle exercise, but the chosen approach 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.

The analysis of zero-crossings has been pioneered by Rice (1944) who derives a link between the autocorrelation (acf) of a zero-mean stationary Gaussian process and its expected number of crossings in a fixed interval. Interestingly, sign changes of 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.

The optimization criterion is derived in section 2.2 with a discussion of robusteness and extensions of the basic methodological framework; solutions of the criterion are proposed in section 2.3 with a discussion of boundary and singular cases, numerical aspects as well as the sample distribution; section 2.5 illustrates various applications including ordinary forecasting, a smoothness-timeliness dilemma, multiplicity and uniqueness features as well as a fully worked-out singular case; finally, section 2.9 concludes by summarizing our main findings.

## 2.2 Simple Sign-Accuracy (SSA-) Criterion

Let $\epsilon_t, t \in \mathbb{Z}$, be Gaussian standard white noise[1] and let $\gamma_k \in \mathbb{R}$ for $k \in \mathbb{Z}$ be a square summable sequence $\sum_{k=-\infty}^{\infty} \gamma_k^2 < \infty$. Then $z_t = \sum_{k=-\infty}^{\infty} \gamma_k \epsilon_{t-k}$ is a stationary Gaussian zero-mean process with variance $\sum_{k=-\infty}^{\infty} \gamma_k^2$. We consider estimation of $z_{t+\delta}$, $\delta \in \mathbb{Z}$, referred to as the *target*, based on the predictor $y_t := \sum_{k=0}^{L-1} b_k \epsilon_{t-k}$, where $b_k$ are the coefficients of a finite-length one-sided causal filter. This problem is commonly referred to as fore-, now- or backcast, depending on $\delta > 0$, $\delta = 0$ or $\delta < 0$. An extension of this framework to $z_t = \sum_{k=-\infty}^{\infty} \gamma_k x_{t-k}$, where $x_t = \sum_{j \geq 0} \xi_j \epsilon_{t-j}$ is a stationary (or non-stationary integrated) process, is proposed in section 2.2.2. However, for clarity of exposition and notational convenience we henceforth assume $x_t = \epsilon_t$, our *basic* methodological framework, acknowledging that straightforward modifications apply in the case of autocorrelated $x_t$.

### 2.2.1 Sign-Accuracy, MSE and Holding-Time

We look for an estimate $y_t$ of $z_{t+\delta}$ such that the probability $\mathrm{P}\Big(\mathrm{sign}(z_{t+\delta}) = \mathrm{sign}(y_t)\Big)$ is maximized as a function of $\mathbf{b} = (b_0, ..., b_{L-1})'$. We now refer to this criterion in terms of *sign accuracy* (SA).

**Proposition 1** *Under the above assumptions about $\epsilon_t, z_t$ the sign accuracy criterion can be stated as*

$$\max_{\mathbf{b}} \rho(y, z, \delta) \tag{2.1}$$

*where*

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

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

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

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

**Remarks**

We here discard the scaling parameter from further consideration since our approach emphasizes signs, smoothness and to some extent timeliness aspects as foremost priorities. In this perspective, predictors that differ by an arbitrary (positive) normalization constant are felt equivalent. Note also that classification methods such as e.g. logit models are less suitable for the purpose at hand because fitting the signs of $z_{t+\delta}$, instead of the actual observations $z_{t+\delta}$, would result in a loss of efficiency under the above assumptions.

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

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

$$ht(y|\mathbf{b}) = \frac{\pi}{\arccos(\rho(y,y,1))} \tag{2.2}$$

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

A proof is provided by Kedem (1986). 'Smoothness' of the predictor $y_t$ is formalized by constraining $\mathbf{b}$ such that

$$ht(y|\mathbf{b}) = ht_1 \tag{2.3}$$

or, equivalently,

$$\rho(y,y,1) = \rho_1 \tag{2.4}$$

where $ht_1$ or $\rho_1$, linked through 2.2, are proper hyper-parameters of our design. In the following, we refer to the 'holding-time' either in terms of $ht(y|\mathbf{b})$ or $\rho(y,y,1)$, clarifying our intent in case of ambiguity. We here argue that the hyper-parameter $ht_1$ is interpretable and can be set 'a priori', at the onset of an analysis, according to structural elements of a prediction problem. Wildi (2023) illustrates the proceeding in a business-cycle application, where $ht_1$ matches the length of historical recession episodes. As an alternative, $ht_1$ could be sized in view of taming 'trading' costs generated by adjustments of market-exposures at sign-changes of the growth-rate (bullish vs. bearish markets). In any case, keeping the rate of unsystematic 'noisy' crossings under control can improve an assessment of the current state of an observed phenomenon, in terms of up- or down-swing phases, see Wildi (2023). The following proposition then derives upper and lower bounds for admissible constraints.

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

We refer to N. Davies, M. B. Pate and M. G. Frost (1974) for a proof, see also proposition 5 further down. Consider now the sign accuracy criterion 2.1 endowed with the holding-time constraint 2.4:

$$\left.\begin{array}{c} \max_{\mathbf{b}} \dfrac{\sum_{k=0}^{L-1}\gamma_{k+\delta}b_k}{\sqrt{\sum_{k=-\infty}^{\infty}\gamma_k^2}\sqrt{\sum_{k=0}^{L-1}b_k^2}} \\ \dfrac{\sum_{k=1}^{L-1}b_{k-1}b_k}{\sum_{k=0}^{L-1}b_k^2} = \rho_1 \end{array}\right\} \tag{2.5}$$

This optimization problem is called *simple sign-accuracy* or SSA-criterion and we refer to solutions of this criterion by the acronym SSA or SSA$(ht_1, \delta)$ or SSA$(\rho_1, \delta)$ to stress the dependence of the predictor on the pair of hyper-parameters, see section 2.5.2 for reference. The SSA-criterion merges MSE, sign accuracy and smoothing requirements in a flexible and consistent way. Departures from the Gaussian assumption can be accommodated in the sense that $y_t$ or $z_t$ can be 'nearly Gaussian' even if $x_t = \epsilon_t$ is not, due to the central limit theorem, see Wildi (2023) for an application to financial data (equity index). Finally, the criterion remains appealing outside of a strict holding-time or zero-crossing perspective by complementing the classic predictor with a generic smoothing constraint.

### 2.2.2 Extension to Stationary Processes

Let

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

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

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

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

$$\left.\begin{array}{c} \max\limits_{(\mathbf{b}\cdot\boldsymbol{\xi})} \dfrac{\sum_{k\geq 0}(\gamma\cdot\xi)_{k+\delta}(b\cdot\xi)_k}{\sqrt{\sum_{|k|<\infty}(\gamma\cdot\xi)_k^2}\sqrt{\sum_{j\geq 0}(b\cdot\xi)_j^2}} \\ \dfrac{\sum_{j\geq 1}(b\cdot\xi)_{j-1}(b\cdot\xi)_j}{\sum_{j\geq 0}(b\cdot\xi)_j^2} = \rho_1 \end{array}\right\} \tag{2.6}$$

which can be solved for $(b\cdot\xi)_j, j = 0, 1, ...$, see theorem 1. The sought-after filter coefficients $b_k$ can then be obtained from $(b\cdot\xi)_j$ by inversion. Note that non-stationary integrated processes could be handled similarly, after suitable differencing. Also, we refer to standard results in textbooks

for a derivation of $\xi_k$ or $\epsilon_t$ based on a finite sample $x_1, ..., x_T$, see e.g. Brockwell and Davis (1993). Finally, for the sake of notational simplicity we henceforth assume $x_t = \epsilon_t$ to be white noise, acknowledging that straightforward modifications would apply in the case of autocorrelation.

## 2.3 Solution of the SSA-Criterion: Frequency-Domain

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

**Proposition 4** *Let* $\hat{z}_{t+\delta} = \sum_{k=0}^{L-1} \gamma_{k+\delta}\epsilon_{t-k} = \boldsymbol{\gamma}'_\delta \boldsymbol{\epsilon}_t$ *denote the classic MSE-predictor. Then* $z_t$ *in criterion 2.5 can be replaced by* $\hat{z}_{t+\delta}$.

Proof

A proof follows from

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

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

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

The proposition suggests that the SSA-predictor $y_t$ should 'fit' the MSE-predictor $\hat{z}_{t+\delta}$ while complying with the holding-time constraint. Therefore, we henceforth refer to $\hat{z}_{t+\delta}$ (or $\boldsymbol{\gamma}_\delta$) as an equivalent target specification. Let then

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

of dimension $L * L$ designate the so-called autocovariance-generating matrix so that $\rho(y, y, 1) = \frac{\mathbf{b'Mb}}{\mathbf{b'b}}$. The following proposition relates stationary points of the lag-one autocorrelation $\rho(y, y, 1)$ with eigenvectors and eigenvalues of $\mathbf{M}$.

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

Proof

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

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

A stationary point of $\rho(y,y,1)$ is found by equating the derivative of the Lagrangian $\mathfrak{L} = \mathbf{b}'\mathbf{M}\mathbf{b} - \lambda(\mathbf{b}'\mathbf{b} - 1)$ to zero i.e.

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

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

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

for some $i \in \{1,...,L\}$ and therefore $\rho(y,y,1)$ must be the corresponding eigenvalue, as claimed. Since the unit-sphere is free of boundary-points we conclude that the extremal values $\rho_{min}(L)$, $\rho_{max}(L)$ must be stationary points i.e. $\rho_{min}(L) = \min_i \lambda_i$ and $\rho_{max}(L) = \max_i \lambda_i$.

We now identify filter coefficients and corresponding filter outputs so that e.g. $y_t$ and $\mathbf{b}$ will both be referred to as predictor or estimate (and similarly for the target(s)). Let then $\lambda_i, \mathbf{v}_i$ denote the pairings of eigenvalues and eigenvectors of $\mathbf{M}$, ordered according to the increasing size of $\lambda_i$ and let $\mathbf{V}$ designate the orthonormal basis of $\mathbb{R}^L$ based on the column-vectors $\mathbf{v}_i$, $i = 1,...,L$. We then consider the spectral decomposition of the target $\boldsymbol{\gamma}_\delta \neq \mathbf{0}$

$$\boldsymbol{\gamma}_\delta = \sum_{i=n}^{m} w_i\mathbf{v}_i = \mathbf{V}\mathbf{w} \tag{2.7}$$

with (spectral-) weights $\mathbf{w} = (w_1,...,w_L)'$, where $1 \leq n \leq m \leq L$ and $w_m \neq 0, w_n \neq 0$. If $n > 1$ or $m < L$ then $\boldsymbol{\gamma}_\delta$ is called *band-limited*. Also, we refer to $\boldsymbol{\gamma}_\delta$ as having *complete* (or *incomplete*) spectral support depending on $w_i \neq 0$ for $i = 1,...,L$ (or not). Finally, denote by $NZ := \{i|w_i \neq 0\}$ the set of indexes of non-vanishing weights $w_i$. The following theorem derives a parametric functional form of the SSA solution under various assumptions about the problem specification.

**Theorem 1** *Consider the SSA optimization problem 2.5 and the following set of regularity assumptions:*

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

*Then*

1. *If the third regularity assumption is violated (admissibility) and if* $|\rho_1| > \rho_{max}(L)$*, then the problem cannot be solved unless the filter-length* $L$ *is increased such that* $|\rho_1| \leq \rho_{max}(L)$*.*

*On the other hand, if* $\rho_1 = \lambda_1 = -\rho_{max}(L)$ *or* $\rho_1 = \lambda_L = \rho_{max}(L)$ *(limiting cases), then* $sign(w_1)\mathbf{v}_1$ *or* $sign(w_L)\mathbf{v}_L$ *are the corresponding solutions of the SSA-criterion (up to arbitrary scaling), where* $w_i$ *are the spectral weights in 2.7 and where it is assumed that* $w_1 \neq 0$*, if* $\rho_1 = \lambda_1$*, or* $w_L \neq 0$*, if* $\rho_1 = \lambda_L$*.*

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

$$\mathbf{b} = D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta = D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \tag{2.8}$$

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

3. *If all regularity assumptions hold, then the lag-one autocorrelation of* $\mathbf{b}$ *in 2.8 is*

$$\rho(\nu) := \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} = \frac{\sum_{i=1}^{L} \lambda_i w_i^2 \frac{1}{(2\lambda_i - \nu)^2}}{\sum_{i=1}^{L} w_i^2 \frac{1}{(2\lambda_i - \nu)^2}} \tag{2.9}$$

*and* $\nu = \nu(\rho_1)$ *can always be selected such that the SSA-solution* $\mathbf{b} = \mathbf{b}(\nu(\rho_1))$ *in 2.8 complies with the holding-time constraint.*

4. *If* $|\nu| > 2\rho_{max}(L)$ *then* $\rho(\nu)$ *as defined in 2.9 is a strictly monotonic function in* $\nu$ *and the parameter* $\nu$ *in 2.8 is determined uniquely by the holding-time constraint* $\rho(\nu) = \rho_1$*.*

Proof

The SSA-problem 2.5 can be rewritten as

$$\begin{aligned} \max_{\mathbf{b}} \boldsymbol{\gamma}_\delta'\mathbf{b} & \\ \mathbf{b}'\mathbf{b} &= 1 \\ \mathbf{b}'\mathbf{M}\mathbf{b} &= \rho_1 \end{aligned} \tag{2.10}$$

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

$$\mathbf{b} := \sum_{i=1}^{L} \alpha_i \mathbf{v}_i \tag{2.11}$$

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

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

so that

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

where $j_0$ is such that $\lambda_{j_0} \neq 0$[2]. The SSA-problem can be solved if the ellipse, defined by the holding-time constraint, intersects the unit-sphere. For this purpose we plug the former equation into the latter:

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

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

$$\alpha_{i_0} = \pm\sqrt{\frac{\lambda_{j_0} - \rho_1}{\lambda_{j_0} - \lambda_{i_0}} - \sum_{k\neq i_0, k\neq j_0} \alpha_k^2 \frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}}} \tag{2.12}$$

Under the case posited in assertion 1 $\rho_1 = \lambda_{i_0}$ with either $i_0 = 1$, i.e. $\rho_1 = -\rho_{max}(L)$, or $i_0 = L$, i.e. $\rho_1 = \rho_{max}(L)$. Let then $i_0 = 1$ so that 2.12 becomes

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

Assume also $j_0 = 2$ (a similar proof can be derived for arbitrary $j_0 \leq 2$, see footnote 3) so that $\lambda_2 - \lambda_k < 0$ in the nominator and $\lambda_2 - \lambda_1 > 0$ in the denominator in the last expression. Therefore, the term under the square-root is larger than one if $\alpha_k \neq 0$ for some $k > 2$ which would imply $|\alpha_1| > 1$ thus contradicting the unit-sphere constraint[3]. We then deduce $\alpha_k = 0$ for $k > 2$ so that $\alpha_1 = \pm 1$ and $\alpha_2 = 0$ and therefore $\pm\mathbf{v}_1$ are the only admissible potential solutions of the SSA-problem (the contacts of ellipse and unit-sphere are tangential). Since $w_1 \neq 0$ by assumption, the solution must be $\mathbf{b} := \text{sign}(w_1)\mathbf{v}_1$ because it maximizes the criterion value $\boldsymbol{\gamma}_\delta'\mathbf{b} = \text{sign}(w_1)w_1 > 0$. If $w_1 = 0$ then the problem is ill-conditioned in the sense that the only possible solutions $\pm\mathbf{v}_1$ do not correlate with the target $z_{t+\delta}$ anymore. Note that similar reasoning applies if $i_0 = L$ (setting $j_0 = L - 1$ in 2.12 and assuming $w_L \neq 0$), which concludes the proof of the first assertion.
To show the second assertion we now assume that all regularity assumptions hold and we define the Lagrangian function

$$L := \boldsymbol{\gamma}_\delta'\mathbf{b} - \lambda_1(\mathbf{b}'\mathbf{b} - 1) - \lambda_2(\mathbf{b}'\mathbf{M}\mathbf{b} - \rho_1) \tag{2.13}$$

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

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

Note that the second regularity assumption (non-degenerate case) implies that the holding-time constraint 2.10 is 'active' i.e. $\lambda_2 \neq 0$. Dividing by $\lambda_2$ then leads to

$$D\boldsymbol{\gamma}_\delta = \boldsymbol{\nu}\mathbf{b} \tag{2.14}$$

$$\boldsymbol{\nu} := (2\mathbf{M} - \nu\mathbf{I}) \tag{2.15}$$

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

[3] Similar contradictions could be derived for any $j_0 > 1$ since $\left|\frac{\lambda_{j_0} - \lambda_k}{\lambda_{j_0} - \lambda_{i_0}}\right| < 1$ if $i_0 = 1$ so that if any $\alpha_k \neq 0$, for $k \neq 1, j_0$, then equation 2.12 would conflict with the unit-sphere constraint.

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

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

where we rely on the spectral decomposition 2.11 of $\mathbf{b}$. By assumption $L \geq 3$ (smoothing) so that $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_L)'$ is defined on a $L - 2 \geq 1$ dimensional intersection of unit-sphere and holding-time constraints. We then infer that the objective function is not overruled by the constraint i.e. $|\lambda_2| < \infty$ so that $D \neq 0$ in 2.14, as claimed. Furthermore, equation 2.14 can be written as

$$\begin{aligned} b_{k+1} - \nu b_k + b_{k_1} &= D\gamma_{k+\delta}\ ,\ 1 \leq k \leq L-2 \\ b_1 - \nu b_0 &= D\gamma_\delta\ ,\ k = 0 \\ -\nu b_{L-1} + b_{L-2} &= D\gamma_{L-1+\delta}\ ,\ k = L-1 \end{aligned} \tag{2.16}$$

for $k = 0, ..., L-1$ so that $b_{-1} = b_L = 0$ are implicitly assumed for the natural extension $(b_{-1}, \mathbf{b}, b_L)'$ of the time-invariant linear filter. The eigenvalues of $\boldsymbol{\nu}$ are $2\lambda_i - \nu$ with corresponding eigenvectors $\mathbf{v}_i$. We note that if $\mathbf{b}$ is the solution of the SSA-problem, then $\nu/2$ cannot be an eigenvalue of $\mathbf{M}$ since otherwise $\boldsymbol{\nu}$ in 2.14 would map one of the eigenvectors in the spectral decomposition of $\mathbf{b}$ to zero which would contradict the last regularity assumption (completeness). Therefore we can assume that $\boldsymbol{\nu}^{-1}$ exists and

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

where the diagonal matrix $\mathbf{D}_\nu^{-1}$ has entries $\frac{1}{2\lambda_i - \nu}$. We can then solve 2.14 for $\mathbf{b}$ and obtain

$$\begin{aligned} \mathbf{b} &= D\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta \qquad (2.17) \\ &= D\mathbf{V}\mathbf{D}_\nu^{-1}\mathbf{V}'\mathbf{V}\mathbf{w} \\ &= D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \qquad (2.18) \end{aligned}$$

where we inserted 2.7. Since $\boldsymbol{\nu}$ has full rank, the solution of the SSA-problem is uniquely determined by $\nu$, at least down to arbitrary scaling, hereby completing the proof of assertion 2.
We next proceed to assertion 3 and consider

$$\begin{aligned} \rho(\nu) &= \rho(y(\nu), y(\nu), 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}} \\ &= \frac{\left(D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)' \mathbf{M} \left(D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)}{\left(D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)' \left(D\sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i\right)} \\ &= \frac{\sum_{i=1}^{L} \frac{\lambda_i w_i^2}{(2\lambda_i - \nu)^2}}{\sum_{i=1}^{L} \frac{w_i^2}{(2\lambda_i - \nu)^2}} \end{aligned} \tag{2.19}$$

where we inserted 2.18 and made use of orthonormality $\mathbf{v}_i'\mathbf{v}_j = \delta_{ij}$. The last expression implies $\lim_{\nu \to 2\lambda_i} \rho(\nu) = \lambda_i$ for all $i = 1, ..., L$. Since $\lambda_1 = -\rho_{max}(L)$ and $\lambda_L = \rho_{max}(L)$, by proposition 5, we infer that lower and upper boundaries $\pm\rho_{max}(L)$ can be reached by $\rho(\nu)$, asymptotically. Continuity of $\rho(\nu)$ and the intermediate-value theorem then imply that any $\rho_1 \in ]-\rho_{max}(L), \rho_{max}(L)[$

is admissible for the holding-time constraint under the posited assumptions.

We now proceed to assertion 4 by showing that the parameter $\nu$ is determined uniquely by $\rho_1$ in the holding-time constraint if $|\nu| > 2\rho_{max}(L)$. Note that all eigenvalues $2\lambda_i - \nu$ of $\boldsymbol{\nu}$ must be (strictly) negative, if $\nu > 2\rho_{max}(L)$, or strictly positive, if $\nu < -2\rho_{max}(L)$, so that all eigenvalues of $\boldsymbol{\nu}^{-1}$, being the reciprocals of the former, must be of the same sign, either all positive or all negative. Finally, the eigenvalues of $\boldsymbol{\nu}$ or $\boldsymbol{\nu}^{-1}$ must be pairwise different since the eigenvalues of $\mathbf{M}$ are so. We then obtain

$$\begin{aligned}
\frac{\partial \rho\Big(y(\nu), y(\nu), 1\Big)}{\partial \nu} &= \frac{\partial}{\partial \nu}\left(\frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}}\right) = \frac{\partial}{\partial \nu}\left(\frac{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\,\prime}\mathbf{M}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-1\,\prime}\boldsymbol{\nu}^{-1}\boldsymbol{\gamma}_\delta}\right) = \frac{\partial}{\partial \nu}\left(\frac{\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}{\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-2}\boldsymbol{\gamma}_\delta}\right) \\
&= \frac{2\boldsymbol{\gamma}_\delta'\mathbf{M}\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta\mathbf{b}'\mathbf{b}/D - (2\mathbf{b}'\mathbf{M}\mathbf{b}/D)\boldsymbol{\gamma}_\delta'\boldsymbol{\nu}^{-3}\boldsymbol{\gamma}_\delta}{((\mathbf{b}'\mathbf{b})^2/D^2)} \\
&= \frac{2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b}/D^2 - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}/D^2}{\mathbf{b}'\mathbf{b}/D^2} \\
&= 2\mathbf{b}'\mathbf{M}\boldsymbol{\nu}^{-1}\mathbf{b}\mathbf{b}'\mathbf{b} - 2\mathbf{b}'\mathbf{M}\mathbf{b}\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b} \qquad (2.20)
\end{aligned}$$

where 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 $\boldsymbol{\nu}^{-1\,\prime} = \boldsymbol{\nu}^{-1}$ (symmetry) and we relied on generic matrix differentiation rules in the third equation[4]; finally we relied on $\mathbf{b}'\mathbf{b} = 1$ in the last equation. We can now insert

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

which is a reformulation of $(2\mathbf{M} - \nu\mathbf{I})\boldsymbol{\nu}^{-1} = \mathbf{I}$ into the first summand in 2.20 to obtain

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

We can now insert this expression into 2.20 and isolate $\mathbf{b}'\boldsymbol{\nu}^{-1}\mathbf{b}$ to obtain

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

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

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

$(w_1, ..., w_L)'$. Therefore

$$\begin{aligned}\frac{\partial\rho\Big(y(\nu),y(\nu),1\Big)}{\partial\nu} &= -\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-3}\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-1} + \left(\sum_{j=0}^{L-1} w_j^2\lambda_{j\nu}^{-2}\right)^2 \\ &= -\sum_{i>k} w_i^2 w_k^2\Big(\lambda_{i\nu}^{-1}\lambda_{k\nu}^{-3} + \lambda_{i\nu}^{-3}\lambda_{k\nu}^{-1} - 2\lambda_{i\nu}^{-2}\lambda_{k\nu}^{-2}\Big) \qquad (2.21)\end{aligned}$$

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 all of the same sign, pairwise different and non-vanishing if $|\nu| > 2\rho_{max}(L)$. Since $w_i \neq 0$ (last regularity assumption: completeness) we deduce $w_i^2 w_k^2 \neq 0$ in 2.21. 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 $\nu$ if $|\nu| > 2\rho_{max}(L)$, as claimed.

**Remark**

Gaussianity is not required in the derivation of the above proof because the SSA-criterion 2.5 relies solely on correlations. The Gaussian hypothesis is needed when establishing formal links between correlations and sign-accuracy or holding-time concepts but $y_t$ or $z_t$ can be nearly Gaussian even if $\epsilon_t$ isn't.

**Corollary 1** *Let all regularity assumptions of the previous theorem hold except completeness so that $NZ \subset \{1, ..., L\}$ or, stated otherwise, there exists $i_0$ such that $w_{i_0} = 0$ in 2.7. Then:*

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

$$\mathbf{b}(\nu) = D\sum_{i\in NZ}\frac{w_i}{2\lambda_i - \nu}\mathbf{v}_i \qquad (2.22)$$

*with corresponding lag-one acf*

$$\rho(\nu) = \frac{\sum_{i\in NZ}\frac{\lambda_i w_i^2}{(2\lambda_i-\nu)^2}}{\sum_{i\in NZ}\frac{w_i^2}{(2\lambda_i-\nu)^2}} =: \frac{M_1}{M_2} \qquad (2.23)$$

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

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

$$\mathbf{b}_{i_0}(\tilde{N}_{i_0}) := \mathbf{b}(\nu_{i_0}) + D\tilde{N}_{i_0}\mathbf{v}_{i_0} \qquad (2.24)$$

*with lag-one acf*

$$\rho_{i_0}(\tilde{N}_{i_0}) = \frac{M_{i_0 1} + \lambda_{i_0}\tilde{N}_{i_0}^2}{M_{i_0 2} + \tilde{N}_{i_0}^2} \qquad (2.25)$$

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

$$\tilde{N}_{i_0} = \pm\sqrt{\frac{\rho_1 M_{i_0 2} - M_{i_0 1}}{\lambda_{i_0} - \rho_1}} \tag{2.26}$$

*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 maximal criterion value.*

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

Proof

The first assertion follows directly from the Lagrangian equation 2.14

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

where $\boldsymbol{\nu}$ has full rank if $\nu \in \mathbb{R} - \{2\lambda_i | i = 1, ..., L\}$, as assumed. Under the case posited in the second assertion $\boldsymbol{\nu}_{i_0}$ does not have full rank anymore and $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ as defined by 2.24 is a solution of the Lagrangian equation

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

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

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

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

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

We infer that $N_{i_0}$ is always positive if $0 < \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} < \rho_1 < \lambda_{i_0}$ or $0 > \rho(\nu_{i_0}) = \frac{M_{i_0 1}}{M_{i_0 2}} > \rho_1 > \lambda_{i_0}$, so that $\tilde{N}_{i_0} = \pm\sqrt{N_{i_0}} \in \mathbb{R}$, as claimed. Finally, the correct sign combination of the pair $D, \tilde{N}_{i_0}$ is determined by the maximal criterion value.
For a proof of the third and last assertion we first assume that $\boldsymbol{\gamma}_\delta$ is not band-limited so that $w_1 \neq 0$ and $w_L \neq 0$. Then, $\lim_{\nu\to 2\lambda_1}\rho(\nu) = \lambda_1 = -\rho_{max}(L)$ and $\lim_{\nu\to 2\lambda_L}\rho(\nu) = \lambda_L = \rho_{max}(L)$, see proposition 5. By continuity of $\rho(\nu)$ and by virtue of the intermediate-value theorem we then infer that any $\rho_1$ 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 2.24, can 'fill the gap' and reach out the lower boundary $-\rho_{max}(L)$ as $\tilde{N}_1 \to \infty$. A similar reasoning would apply in the case $w_L = 0$ which achieves the proof of the corollary.

The case of incomplete spectral support is illustrated by a worked-out example in section 2.5.4. In order to simplify exposition, we now assume that $|\nu| > 2$. In fact, $|\nu| \leq 2$ would imply that the

solution of the homogeneous difference-equation

$$b_{k+1} - \nu b_k + b_{k_1} \quad = \quad 0$$

would be subject to a unit-root so that the coefficients would not decay to zero for increasing lag. Since the SSA-solution specified by 2.16 is obtained by a suitable linear combination of non-homogeneous and homogeneous solutions[5], we then infer that its coefficients would not decay to zero either with increasing lag, hence suggesting evidence of an ill-posed prediction problem. Typically, this issue could be addressed by selecting $L$ sufficiently large i.e. at least twice the imposed holding-time.

**Corollary 2** *Let the assumptions of theorem 1 hold and assume $|\nu| > 2$. Then the solution to the SSA-optimization problem 2.5 is*

$$\mathbf{b}(\nu_0) = sign_{\nu_0} \sum_{i=1}^{L} \frac{w_i}{2\lambda_i - \nu_0} \mathbf{v}_i \tag{2.27}$$

*where $\nu_0$ is the unique solution to the non-linear equation*

$$\frac{\mathbf{b}(\nu_\mathbf{0})' \mathbf{M} \mathbf{b}(\nu_\mathbf{0})}{\mathbf{b}(\nu_\mathbf{0})' \mathbf{b}(\nu_\mathbf{0})} = \rho_1 \tag{2.28}$$

*The sign $sign_{\nu_0} = \pm 1$ is selected such that $\mathbf{b}(\nu_\mathbf{0})' \boldsymbol{\gamma}_\delta > 0$ (positive criterion value).*

A proof follows readily from assertions 2-4 of theorem 1, noting that $|\nu| > 2 > 2\rho_{max}(L)$. In this case, numerical computations are swift due to strict monotonicity and also because 2.27 does not rely on a matrix inversion, unlike 2.17. If $|\nu| \leq 2$ then strict monotonicity and uniqueness of the solution of 2.28 are lost, see section 2.5.3 for a worked-out example. To conclude, the following corollary derives the distribution of the SSA-predictor.

**Corollary 3** *Let all regularity assumptions of theorem 1 hold and let $\hat{\boldsymbol{\gamma}}_\delta$ be a finite-sample estimate of the MSE-predictor $\boldsymbol{\gamma}_\delta$ with mean $\boldsymbol{\mu}_{\gamma_\delta}$ and variance $\boldsymbol{\Sigma}_{\gamma_\delta}$. Then mean and variance of the SSA-predictor $\hat{\mathbf{b}}$ are*

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

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

The proof readily follows from 2.8. Note that mean, variance and (asymptotic) distribution of the MSE-estimate under various assumptions about $x_t$ are derived in standard textbooks, see e.g. Brockwell and Davis (1993).

[5]The second regularity assumption of the theorem (non-degenerate case) implies that the weight assigned to the homogeneous solution is non-vanishing.

## 2.4 Solution of the SSA-Problem: Time-Domain

We here consider the AR(2) difference-equation 2.16 together with the boundary constraints $b_{-1} = b_L = 0$ as determinants of the SSA-predictor in the time-domain. In this context, the roots of the characteristic AR(2)-polynomial of 2.16 are denoted by $\lambda_{1\rho_1}$ and $\lambda_{2\rho_1}$. If $|\nu| < 2$ then the roots are complex conjugate and $\nu = 2\Re(\lambda_{1\rho_1})$. If $|\nu| \geq 2$ then the roots are real inverse i.e. $\lambda_{2\rho_1} = 1/\lambda_{1\rho 1}$ and $\nu = \lambda_{1\rho 1} + 1/\lambda_{1\rho 1}$. Section 2.4.1 proposes a solution of the SSA-problem for a target $z_t$ following an ARMA-specification; section 2.4.2 derives a closed-form solution of the SSA-predictor in a special case.

### 2.4.1 ARMA-Target

**Corollary 4** *Let the regularity assumptions of theorem 1 hold, let $|\nu| > 2$ in 2.16 and let $\lambda_{1\rho_1}$ designate the (real) stable root in $\nu = \lambda_{1\rho_1} + 1/\lambda_{1\rho_1}$. Assume, also, that $\boldsymbol{\gamma}_\delta$ follows an ARMA$(p,q)$-specification with AR- and MA-parameters $a_1, ..., a_p$ and $b_1, ..., b_q$. Then the solution of the SSA-criterion is*

$$b_k := \left\{ \left( \left( \mathbf{A}^{k+1+\delta} - \mathbf{C}_1 \lambda_{1\rho_1}^{k+1} - \mathbf{C}_2 \lambda_{1\rho_1}^{L-k} \right) (\mathbf{A}^2 - \nu \mathbf{A} + \mathbf{I})^{-1} \right) \mathbf{b} \right\}_1 \tag{2.29}$$

*where*

$$\begin{aligned} \mathbf{A} &= \begin{pmatrix} a_1 & a_2 & ... & a_{r-1} & a_r \\ 1 & 0 & ... & 0 & 0 \\ ... & & & & \\ 0 & 0 & ... & 1 & 0 \end{pmatrix} \\ \mathbf{C}_1 &= \frac{1}{1 - \lambda_{1\rho_1}^{2(L+1)}} (\mathbf{A}^\delta - \lambda_{1\rho_1}^{L+1} \mathbf{A}^{L+1+\delta}) \\ \mathbf{C}_2 &= \frac{1}{1 - \lambda_{1\rho_1}^{2(L+1)}} (\mathbf{A}^{L+1+\delta} - \lambda_{1\rho_1}^{L+1} \mathbf{A}^\delta) \end{aligned}$$

$\mathbf{b}' = (1, b_1, ..., b_r)$, $r = \max(p, q+1)$ *and the notation $\{\cdot\}_1$ in 2.29 designates the first component of the corresponding vector.*

Proof

Consider first the MA-inversion

$$\gamma_k = \left( \mathbf{A}^k \mathbf{b} \right)_1$$

of the ARMA-process $z_t$, where $\mathbf{A} = \begin{pmatrix} a_1 & a_2 & ... & a_{r-1} & a_r \\ 1 & 0 & ... & 0 & 0 \\ ... & & & & \\ 0 & 0 & ... & 1 & 0 \end{pmatrix}$, $\mathbf{b}' = (1, b_1, ..., b_r)$ and $r = \max(p, q+1)$. Since

$$\mathbf{A}^{k+2+\delta} - \nu \mathbf{A}^{k+1+\delta} + \mathbf{A}^{k+\delta} = \mathbf{A}^{k+\delta} (\mathbf{A}^2 - \nu \mathbf{A} + \mathbf{I})$$

we deduce that $b'_k := \left\{\left(\mathbf{A}^{k+1+\delta}(\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I})^{-1}\right)\mathbf{b}\right\}_1$ must be a solution of the AR(2) difference equation 2.16, up to arbitrary scaling. Note that $\nu$ must be such that $\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I}$ has full rank, by virtue of the theorem, assuming all regularity assumptions to hold. If $|\nu| > 2$ then the roots of the characteristic AR(2) polynomial 2.16 are real inverse numbers i.e. $\nu = \lambda_{1\rho_1} + 1/\lambda_{1\rho_1}$ where $\lambda_{1\rho_1} \in ]-1, 1[/\{0\}$[6] designates the stable root. Moreover, $\tilde{\mathbf{C}}(k) := \tilde{\mathbf{C}}_1\lambda_{1\rho_1}^{k+1} + \tilde{\mathbf{C}}_2\lambda_{1\rho_1}^{L-k}$, where $\tilde{\mathbf{C}}_1, \tilde{\mathbf{C}}_2$ are arbitrary matrices of dimension $r$, is a solution of the homogeneous difference equation

$$\tilde{\mathbf{C}}(k+1) - \nu\tilde{\mathbf{C}}(k) + \tilde{\mathbf{C}}(k-1) = \mathbf{0}$$

by definition of $\lambda_{1\rho_1}$. Therefore,

$$b_k := \left\{\left(\mathbf{A}^{k+1+\delta}(\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I})^{-1} - \tilde{\mathbf{C}}(k)\right)\mathbf{b}\right\}_1$$

is also a solution of 2.16. We can now select $\tilde{\mathbf{C}}_1$ and $\tilde{\mathbf{C}}_2$ in $\tilde{\mathbf{C}}(k)$ such that the boundary constraints $b_{-1} = b_L = 0$ hold. Specifically, let $\mathbf{C}_i := \tilde{\mathbf{C}}_i(\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I})$, $i = 1, 2$ so that

$$b_k = \left\{\left(\left(\mathbf{A}^{k+1+\delta} - \mathbf{C}_1\lambda_{1\rho_1}^{k+1} - \mathbf{C}_2\lambda_{1\rho_1}^{L-k}\right)(\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I})^{-1}\right)\mathbf{b}\right\}_1 \tag{2.30}$$

We now determine $\mathbf{C}_1, \mathbf{C}_2$ according to the boundary constraints at lags $k = -1$ and $k = L$ by

$$\begin{aligned} \mathbf{A}^{\delta} - \mathbf{C}_1 - \mathbf{C}_2\lambda_{1\rho_1}^{L+1} &= \mathbf{0} \\ \mathbf{A}^{L+1+\delta} - \mathbf{C}_1\lambda_{1\rho_1}^{L+1} - \mathbf{C}_2 &= \mathbf{0} \end{aligned}$$

Solving for $\mathbf{C}_1, \mathbf{C}_2$ leads to

$$\begin{aligned} \mathbf{C}_1 &= \frac{1}{1 - \lambda_{1\rho_1}^{2(L+1)}}(\mathbf{A}^{\delta} - \lambda_{1\rho_1}^{L+1}\mathbf{A}^{L+1+\delta}) \\ \mathbf{C}_2 &= \frac{1}{1 - \lambda_{1\rho_1}^{2(L+1)}}(\mathbf{A}^{L+1+\delta} - \lambda_{1\rho_1}^{L+1}\mathbf{A}^{\delta}) \end{aligned}$$

Inserting these expressions into 2.30 implies that the resulting $b_k$ must be a solution of 2.16 satisfying the implicit boundary constraints $b_{-1} = b_L = 0$ and therefore it must be the SSA-solution, up to arbitrary scaling, by virtue of theorem 1.

**Remarks**

The theorem assumes that $\boldsymbol{\gamma}_\delta$ has complete spectral support. This is always the case for a stationary invertible ARMA-process whose spectral density is strictly positive. But non-invertible processes are allowed, too, assuming that the spectral density does not vanish at a frequency corresponding to an eigenvector $\mathbf{v}_i$, $i = 1, ..., L$, of $\mathbf{M}$: in this case, all spectral weights $w_i$, $i = 1, ..., L$ in 2.7 are non-vanishing, as assumed. Note also that theorem 1 asserts existence of the SSA-solution in the functional form derived in the above corollary and therefore $\nu$ must be such that $\mathbf{A}^2 - \nu\mathbf{A} + \mathbf{I}$ has full rank in the above derivation. Finally, in typical applications the unstable root $1/\lambda_{1\rho_1}$ or,

[6] The singular case $\lambda_{1\rho_1} = 0$ would correspond to the degenerate case $\mathbf{b} \propto \boldsymbol{\gamma}_\delta$ i.e. the SSA-estimate is also the MSA-estimate, which is excluded by the theorem.

equivalently, $\lambda_{1\rho_1}^{L-k}$ in 2.30 can be neglected because $\mathbf{C}_2 = \frac{1}{1-\lambda_{1\rho_1}^{2(L+1)}}(\mathbf{A}^{L+1+\delta} - \lambda_{1\rho_1}^{L+1}\mathbf{A}^{\delta})$ asymptotically vanishes for $L$ sufficiently large. The unit-root case $|\nu| \leq 2$ would behave differently: in particular $b_k$ would not decay to zero anymore with increasing lag $k$, suggesting an ill-posed or 'atypical' prediction problem.

### 2.4.2 Special Case AR(1)-Target and a Closed-Form Solution

We now suggest that the unknown parameter $\lambda_{1\rho_1}$ and therefore $\nu = \lambda_{1\rho_1} + 1/\lambda_{1\rho_1}$ can be obtained in closed-form, without numerical optimization, under certain conditions. To simplify exposition we fix attention to a particular target $z_t$, namely an AR(1)-process, with weights $\gamma_k = \lambda^k$. Extensions to an AR(2)-target are available but for ease of exposition we here omit a corresponding more convoluted derivation, especially since the basic proceeding would remain the same.

**Corollary 5** *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)*
3. *$\lambda_{1\rho_1}, \lambda$ 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)*

$$\lambda_{1\rho_1} = -\frac{1}{3c_3}\left(c_2 + C + \frac{\Delta_0}{C}\right) \tag{2.31}$$

*where*

$$\begin{aligned} C &= \sqrt[3]{\frac{\Delta_1 + sign(\Delta_1)\sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}} \\ \Delta_0 &= c_2^2 - 3c_3c_1 \\ \Delta_1 &= 2c_2^3 - 9c_3c_2c_1 + 27c_3^2c_0 \end{aligned} \tag{2.32}$$

*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 $\lambda$, the forecast horizon $\delta$ and the holding-time constraint $\rho_1$ 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 2.8, down to the correct sign which leads to a positive criterion value $\mathbf{b}'\boldsymbol{\gamma}_\delta \geq 0$.*

Proof

Let $\gamma_{k+\delta} = \lambda^{k+\delta}$. Then

$$b'_k := D\frac{\lambda^\delta}{\lambda^2 - \nu\lambda + 1}\lambda^{k+1} \propto \lambda^{k+\delta} \tag{2.33}$$

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 $\lambda \neq \lambda_{1\rho_1}$, which is always the case by virtue of theorem 1. According to corollary 4 vanishing boundaries $b_{-1} = b_L = 0$ can be imposed by combining $b'_k$ in 2.33 with a suitably scaled solution of the homogeneous difference-equation:

$$b_k = b_k(\lambda_{1\rho_1}) \propto \lambda^{k+\delta} + C_1\lambda_{1\rho_1}^k + C_2\lambda_{1\rho_1}^{L-k} \approx \lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k \tag{2.34}$$

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 4 states also that the unknown stable root $\lambda_{1\rho_1}$ is determined uniquely by requiring

$$\frac{\sum_{k=1}^{L-1} b_k(\lambda_{1\rho_1})b_{k-1}(\lambda_{1\rho_1})}{\sum_{k=0}^{L-1} b_k(\lambda_{1\rho_1})^2} = \rho_1 \tag{2.35}$$

We can now insert 2.34 into this equation and solve for $\lambda_{1\rho_1}$. Specifically, the nominator becomes

$$\sum_{k=1}^{L-1} b_k b_{k-1} \quad = \quad \sum_{k=1}^{L-1}\left(\lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k\right)\left(\lambda^{k-1+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^{k-1}\right) \tag{2.36}$$

The first cross-product of terms in parantheses is

$$\lambda^{-1}\sum_{k=1}^{L-1}\lambda^{2(k+\delta)} = \lambda^{1+2\delta}\sum_{k=0}^{L-2}\lambda^{2k} = \lambda^{1+2\delta}\frac{1-\lambda^{2(L-1)}}{1-\lambda^2} \approx \frac{\lambda^{1+2\delta}}{1-\lambda^2} \tag{2.37}$$

The second cross-product of terms in parentheses is

$$-\lambda_{1\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k+\delta}\lambda_{1\rho_1}^{k-1} = -\lambda_{1\rho_1}\lambda^{2\delta}\sum_{k=0}^{L-2}(\lambda\lambda_{1\rho_1})^k = -\lambda_{1\rho_1}\lambda^{2\delta}\frac{1-(\lambda\lambda_{1\rho_1})^{L-1}}{1-\lambda\lambda_{1\rho_1}} \approx \frac{-\lambda_{1\rho_1}\lambda^{2\delta}}{1-\lambda\lambda_{1\rho_1}} \tag{2.38}$$

The third cross-product of terms in parentheses is

$$-\lambda_{1\rho_1}\lambda^{-1+\delta}\sum_{k=1}^{L-1}\lambda^{k-1+\delta}\lambda_{1\rho_1}^k \approx \frac{-\lambda_{1\rho_1}^2\lambda^{2\delta-1}}{1-\lambda\lambda_{1\rho_1}} \tag{2.39}$$

The last cross-product of terms in parantheses is

$$\lambda_{1\rho_1}^2\lambda^{-2+2\delta}\sum_{k=1}^{L-1}\lambda_{1\rho_1}^{2k-1} = \lambda_{1\rho_1}^3\lambda^{-2+2\delta}\sum_{k=0}^{L-2}\lambda_{1\rho_1}^{2k} = \lambda_{1\rho_1}^3\lambda^{-2+2\delta}\frac{1-\lambda_{1\rho_1}^{2(L-1)}}{1-\lambda_{1\rho_1}^2} \approx \frac{\lambda_{1\rho_1}^3\lambda^{-2+2\delta}}{1-\lambda_{1\rho_1}^2} \tag{2.40}$$

Note that the last assumption of the corollary is critical for the validation of the above approximations. Further, the common denominator of 2.37, 2.38, 2.39 and 2.40 is

$$(1-\lambda^2)(1-\lambda\lambda_{1\rho_1})(1-\lambda_{1\rho_1}^2) \tag{2.41}$$

Summing all terms in 2.37, 2.38, 2.39 and 2.40 under the common denominator 2.41 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 2.35:

$$\sum_{k=0}^{L-1} b_k^2 = \sum_{k=0}^{L-1} \left(\lambda^{k+\delta} - \lambda_{1\rho_1}\lambda^{-1+\delta}\lambda_{1\rho_1}^k\right)^2 \tag{2.42}$$

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 2.41 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 2.41, equation 2.35 can then be re-written as

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

or

$$f_3(\lambda_{1\rho_1}, \rho_1) = 0 \tag{2.43}$$

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[7], see e.g. Cardano's formula.

## 2.5 Examples

Our examples address specific methodological features of the SSA-predictor: a simple introductory forecast exercise is proposed in section 2.5.1; section 2.5.2 presents a more generic prediction problem, emphasizing a smoothness-timeliness dilemma; section 2.5.3 highlights multiplicity and uniqueness results; finally, the singular case of a target with incomplete spectral support is illustrated in section 2.5.4.

### 2.5.1 Example 1: Forecasting

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

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

where $\gamma_k = 1, k = 0, 1, 2$ and with forecast horizon $\delta = 1$ (one-step ahead). For comparison purposes we compute three different SSA-forecast filters $y_{ti}, i = 1, 2, 3$ for $z_t$: the first two are of identical length $L = 20$ with dissimilar holding-times $ht$ =3.74 and 10; the third filter deviates from the second one by selecting $L = 50$; the holding-time of the first filter matches the lag-one autocorrelation of $z_t$ and is obtained by inserting $\rho(z, z, 1) = 2/3$ into 2.2. 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. 2.1 (an arbitrary scaling scheme is applied to SSA filters). Note that predictors based on the 'true' MA(2)-model of $z_t$ are virtually indistinguishable from predictors based on a fitted empirical model, see also table 2.2 below.

[7] 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.

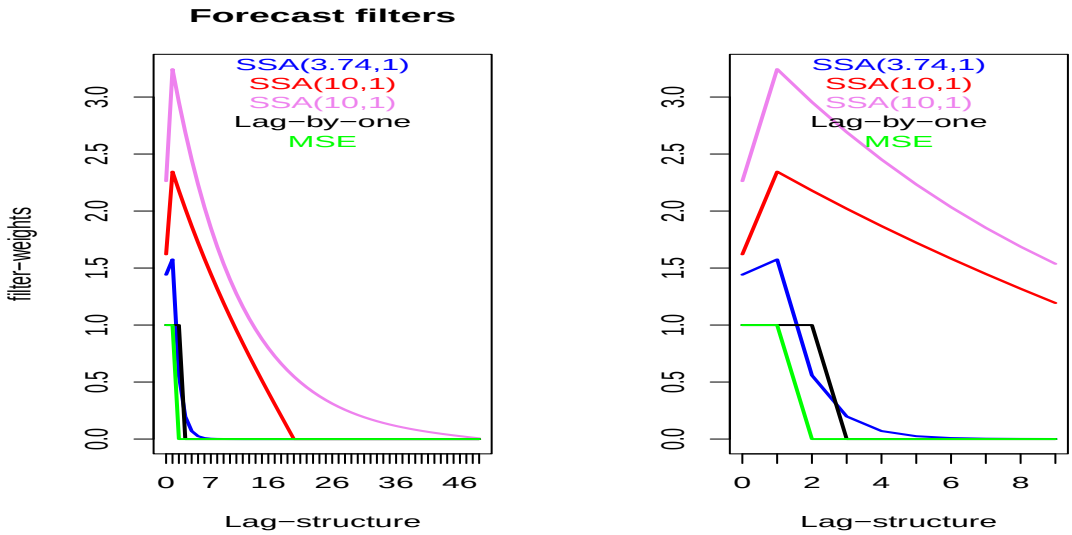


Figure 2.1: Coefficients of MSE-, SSA- and lag-by-one forecast filters with arbitrarily scaled SSA-designs. All lags (left panel) and first ten lags (right panel).

Except for the MSE (green) all other filters rely on past $\epsilon_{t-k}$ for $k > q = 2$ which are required for compliance with the holding-time constraint (stronger smoothing). For a fixed filter-length $L$, a larger holding-time $ht$ asks for a slower zero-decay of filter coefficients (blue vs. red lines) and for fixed holding-time $ht$, a larger $L$ leads to a faster zero-decay but a long tail of the filter (red vs. violet lines). The distinguishing tips of the SSA-filters at lag one in this example are indicative of one of the two implicit boundary constraints, namely $b_{-1} = 0$, see theorem 1. Note that the 'lag-by-one' forecast (black) has the same holding time as the first SSA-filter (blue) so that the latter should outperform the former in terms of sign accuracy or, equivalently, in terms of correlation with the shifted target, as confirmed in table 2.1. MSE outperforms all other forecasts in terms

| | SSA(3.74,1) | SSA(10,1) | SSA(10,1) | Lag-by-one | MSE |
|---|---|---|---|---|---|
| Correlation with target | 0.786 | 0.385 | 0.389 | 0.667 | 0.816 |
| Empirical holding-times | 3.735 | 10.031 | 9.974 | 3.735 | 3.000 |
| Empirical sign accuracy | 0.788 | 0.626 | 0.627 | 0.732 | 0.804 |

Table 2.1: Performances of MSE and lag-by-one benchmarks vs. SSA: All filters are applied to a sample of length 1000000 of Gaussian noise. Empirical holding-times are obtained by dividing the sample-length by the number of zero-crossings.

of correlation and sign accuracy but it loses in terms of smoothness or holding-time; SSA(3.74,1) outperforms the lag-by-one benchmark; both SSA(10,1) loose in terms of sign-accuracy but win in terms of smoothness and while the profiles of longer and shorter filters differ in figure 2.1, their respective performances are virtually indistinguishable in table 2.1, suggesting that the selection of $L$ is not critical (assuming it is at least twice the holding-time). The table also illustrates the

tradeoff between MSE- or sign-accuracy performances of optimal designs, in the top and bottom rows, and smoothing-performances in the middle row (an explicit formal link can be obtained but is omitted here). Finally, table 2.2 displays results when all predictors rely on an empirical model fitted to $z_t$ on a data-sample of length 50: a comparison of both tables suggests that performances are virtually unaffected by the additional estimation step.

| | SSA(3.88,1) | SSA(10,1) | SSA(10,1) | Lag-by-one | MSE |
|---|---|---|---|---|---|
| Correlation with target | 0.774 | 0.385 | 0.387 | 0.680 | 0.815 |
| Empirical holding-times | 3.886 | 10.014 | 10.017 | 3.885 | 2.987 |
| Empirical sign accuracy | 0.782 | 0.626 | 0.627 | 0.738 | 0.803 |

Table 2.2: Same as previous table but all predictors are based on an empirical model of the MA(2)-process fitted in a sample length 50.

### 2.5.2 Example 2: Two Hyperparameters and the Smoothness-Timeliness Dilemma

Often, stronger noise-rejection or smoothing by a (nowcast or forecast) filter is associated with increased lag or 'right-shift' of its output: the following example illustrates that the mentioned tradeoff, a so-called smoothness-timeliness dilemma, does not hold in general. For illustration, we rely on a simple empirical framework where the target is an equally-weighted MA-filter of length $L$ =100 applied to simulated Gaussian noise $\epsilon_t$: $z_t = \frac{1}{100}\sum_{k=0}^{99}\epsilon_{t-k}$. The target must be forecasted at the horizon 20 by a classic MSE as well as a SSA(30,20)-filter, whose holding-time $ht = 30$ exceeds that of the MSE design $ht = 19.8$ by a safe margin. Out of curiosity, we also supply a second SSA(30,40)-filter optimized for forecast horizon 40: the two hyperparameters $ht, \delta$ of the two SSA-designs suggest that for an identical smoothing capability or holding-time, the second filter should have improved timeliness properties in terms of a lead or left-shift. The three (arbitrarily scaled) forecast filters are displayed in fig.2.2[8] and filter outputs, arbitrarily scaled to unit-variance, are compared in fig.2.3.

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

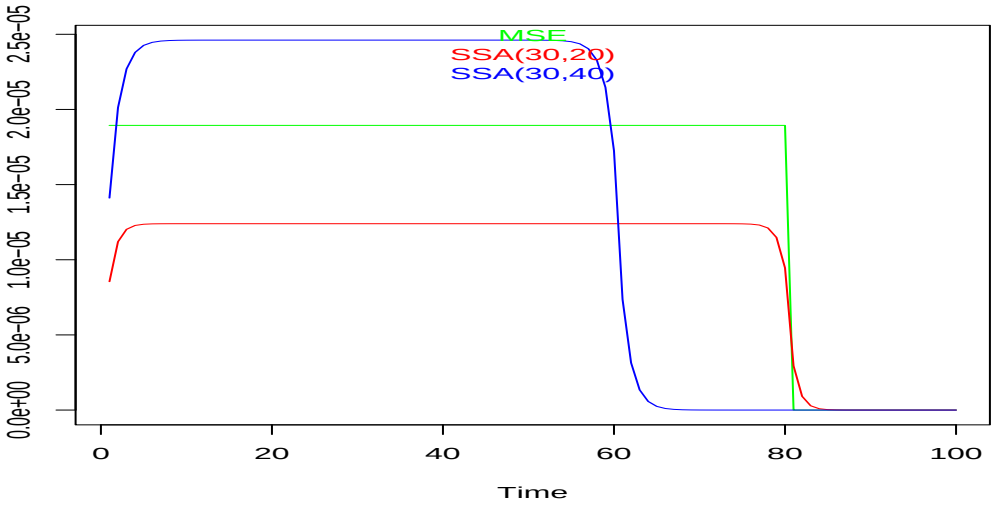


Figure 2.2: Forecast filters: MSE (green), SSA(30,20) (red) and SSA(30,40) (blue) with arbitrary scaling

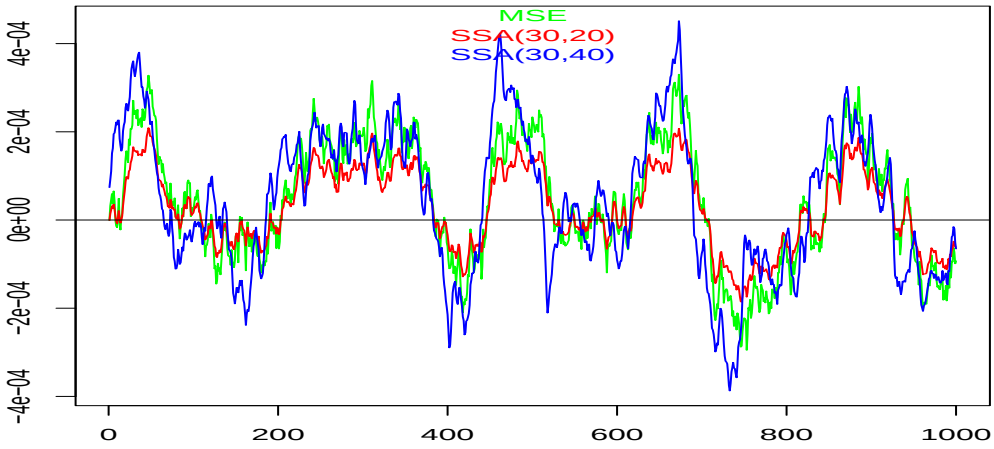


Figure 2.3: Outputs of forecast filters: MSE (green), SSA(30,20) (red) and SSA(30,40) (blue)

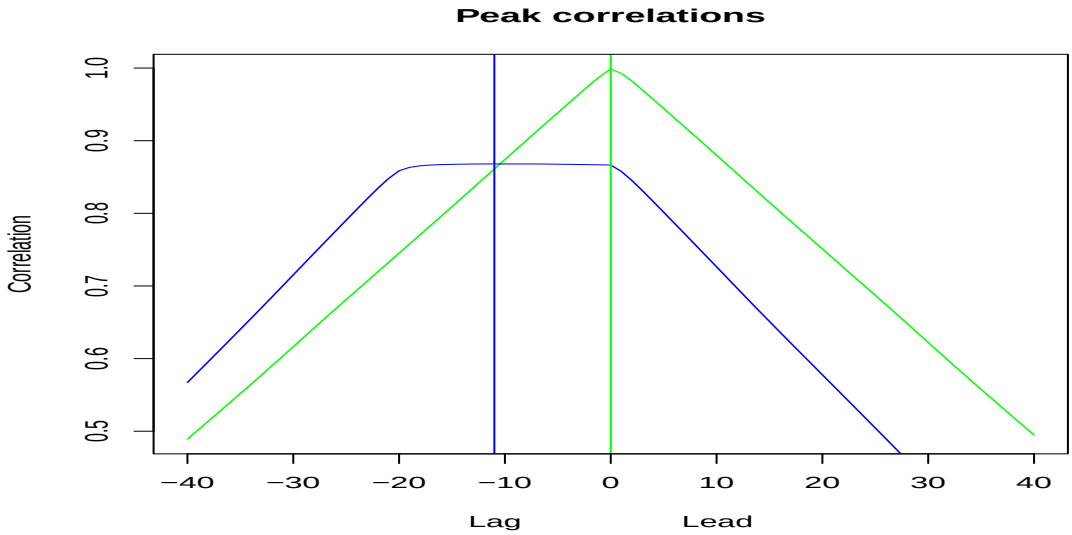


Figure 2.4: Correlation of shifted SSA(30,20) vs. MSE (green) and SSA(30,40) (blue). Positive numbers correspond to a relative lead of SSA(30,20) over the contenders. Peak correlations are indicated by vertical lines.

As expected, the output of SSA(30,40) (blue line in fig.2.3) appears left-shifted. Fig.2.4 displays cross-correlations at various leads and lags of the reference SSA(30,20): the relative shift can be inferred from the peak-correlation i.e. the lead or lag at which the maximum is achieved. The figure suggests that SSA(30,20) and MSE are on par (green line) and that SSA(30,20) lags or, equivalently, that SSA(30,40) leads by 11 time-units (blue line). Finally, the empirical holding-times in table 2.3, computed on a sample of length 100000, conform to expected values, as based on 2.2. We conclude that for identical smoothing capabilities, SSA(30,40) has improved timeliness

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

Table 2.3: Empirical holding-times of MSE and SSA designs

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

### 2.5.3 Example 3: Monotonicity vs. Non-Monotonicity of the Lag-one ACF

We here illustrate uniqueness or multiplicity of the solution of the non-linear holding-time equation 2.28, depending on $|\nu| > 2\rho_{max}(L)$, see assertion 4 of theorem 1. Fig. 2.5 displays the lag-one autocorrelation $\rho(\nu)$ in 2.9 for a SSA-nowcast ($\delta = 0$) as a function of $\nu$ for two different AR(1)-targets $\boldsymbol{\gamma}_0(a_1) = (1, a_1, ..., a_1^9)'$ of length $L = 10$ with $a_1 = 0.99$ (bottom panels) and $a_1 = 0.6$ (top panels). The panels on the left correspond to $|\nu| < 2\rho_{max}(10)$ and illustrate non-monotonicity of $\rho(\nu)$; the panels on the right correspond to $\nu > 2\rho_{max}(10)$ and illustrate strict monotonicity[9].

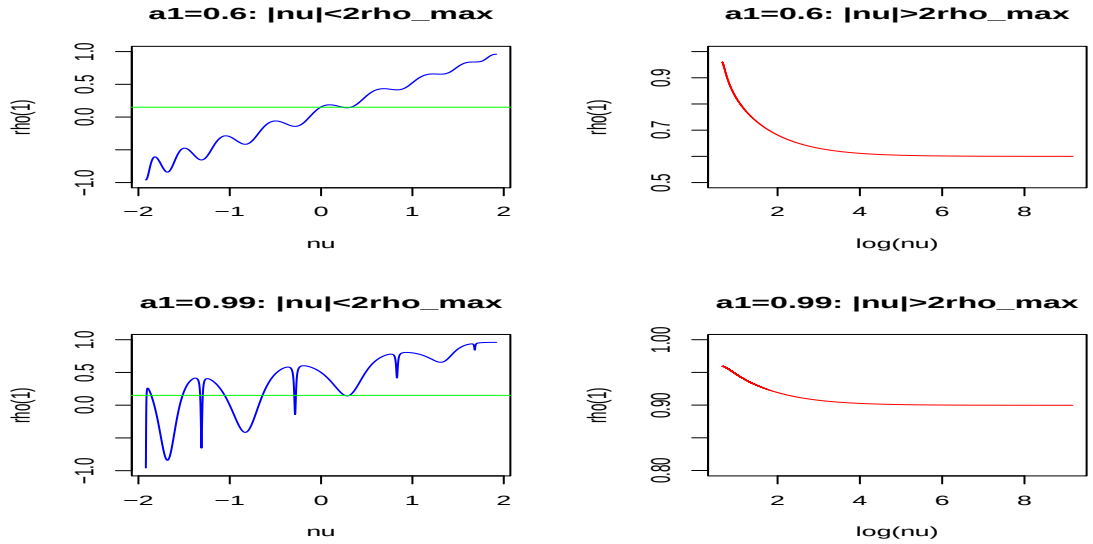


Figure 2.5: Lag-one autocorrelation as a function of nu when the target is a classic AR(1) with a1=0.6 (top) and a1=0.99 (bottom): the left/right-split of the panels corresponds to |nu|<2 rhomax (left) and nu>2 rhomax (right)

Non-monotonicity generally leads to multiple solutions of $\nu$ for given $\rho_1$ for the holding-time equation 2.28, whereby the multiplicity generally depends on $\rho_1$, $L$ as well as on the target $\gamma_{k+\delta}$: as can be seen the green horizontal line corresponding to $\rho_1 = 0.15$ in fig.2.5 intersects the acf four times in the upper (left) panel and $L + 1 = 11$ times in the bottom (left) panel. Monotonicity, on the other hand, means that $\nu$ is determined uniquely by $\rho_1$.

### 2.5.4 Example 4: Application to a Target with Incomplete Spectral Support

In order to illustrate the case of incomplete spectral support, handled by corollary 1, we here consider a simple nowcast example (forecast horizon $\delta = 0$) based on a band-limited $\boldsymbol{\gamma}_0$ of length $L = 10$ whose last three weights in the spectral decomposition 2.7 vanish, $w_8 = w_9 = w_{10} = 0$ ($m = 7$ in 2.7), and the first seven weights are constant $w_i = 0.378$, $i = 1, ..., 7$, whereby

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

$\sum_{i=1}^{10} w_i^2 = 1$. The left panel in fig. 2.6 displays the lag-one acf 2.9 of $\mathbf{b}$ given by 2.8 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 2.25 of the extension $\mathbf{b}_{i_0}(\tilde{N}_{i_0})$ in 2.24, 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 2.25, 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).

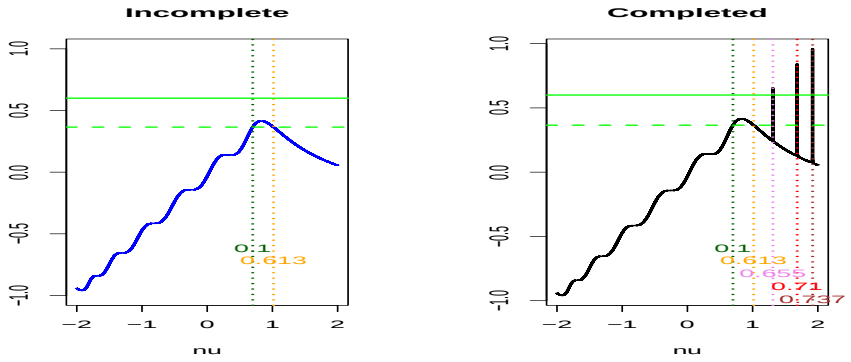


Figure 2.6: Lag-one autocorrelation as a function of nu. Original (incomplete) solutions (left panel) vs. completed solutions (right-panel). Intersections of the acf with the two green lines are potential solutions of the SSA-problem for the corresponding holding-times: criterion values are reported 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, $\mathbf{b}$ 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 2.26, with the correct sign in place.

## 2.6 Application to BCA

We here apply the SSA-design to monthly industrial production indices of various countries with long business-cycle histories and benchmark performances against the HP-filter, see Hodrick and Prescott (1997). We refer to the two-sided symmetric HP-filter as the 'target' $z_t$ which must be nowcasted at the current sample-end i.e. $\delta_0 = 0$. For that purpose, we consider five concurrent designs, namely two classic HP-filters, HP-gap and HP-trend, as well as three SSA-designs based on distinct hyper-parameter settings for $\rho_1, \delta$, see section 2.6.1 for reference. Our implementation of SSA in this application emphasizes simplicity and robustness: in particular, we do not fit models to the data, assuming log-returns to be white noise and accepting the deliberate misspecification as a tradeoff for simplicity; the same filters are applied to all countries; moreover, SSA must at least equal the benchmarks in terms of smoothness or noise-rejection i.e. we emphasize reliability over timeliness. Summing-up, our implementation illustrates the possibility of modifying an existing benchmark in view of emphasizing alternative research- or user-priorities.

### 2.6.1 SSA- and Hodrick Prescott Filters

The HP filter is widely used to estimate trends and cycles of economic time series. It can be interpreted as an optimal MSE-signal extraction filter for the trend in the smooth trend model, see Harvey (1989). Conceptually, this results in 'implied' models for the cycle and the trend, such that applying the HP filter results in MSE optimal estimates. In this framework, the bi-infinite symmetric expansion of the filter is obtained as

$$(\gamma_{|k|}B^k)_{|k|<\infty} = \frac{1}{1+\lambda(1-B)^2(1-B^{-1})^2} \tag{2.44}$$

where $\lambda$ is a 'smoothing' hyperparameter and where $B, B^{-1}$ are backward and forward operators. The implicit data-generating process is an ARIMA(0,2,2) whose MA-coefficients are determined by $\lambda$, see McElroy (2006). In finite samples, the filter behaves differently in the middle or toward the boundaries of the data, where the symmetry is lost: an exact finite sample representation of the concurrent HP-trend filter, denoted as $b_k^{HP-trend}$, is derived in McElroy (2006). In addition, we also consider the classic HP-gap filter

$$b_k^{gap} := \begin{cases} 1 - b_0^{HP-trend} & k = 0 \\ -b_k^{HP-trend} & , \quad k > 0 \end{cases} \tag{2.45}$$

While SSA- and HP-trend will be applied to *differenced* data, the HP-gap is typically applied to data in levels. Therefore, we here propose a modified gap-design $b_k^{\Delta gap}$ such that

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

where $\Delta x_t$ are first differences of a time series $x_t$. Note that filter outputs of original and modified gap-filters are strictly identical but their input series differ. One can verify that $b_k^{\Delta gap} = \sum_{i=1}^{k} b_i^{gap}$ and that the coefficients decay towards zero for increasing lag, see for example McElroy and Wildi

(2020)[10].

According to Morten and Uhlig (2002), we select $\lambda = 14400$ (monthly data). For the holding-time, we select either $ht_1 = 12$, which matches roughly the mean-duration of recessions (see also the closing discussion in section 2.6.2) or $ht_1$ =7.66 which is the expected holding-time of the benchmark HP(trend)-filter, as based on 2.2. Timeliness is addressed by selecting either $\delta = 0$ (nowcast) or $\delta = 18$ (forecast), see fig.2.9 for further analysis and keep in mind that the proper target is a nowcast i.e. $\delta_0 = 0$ is fixed.

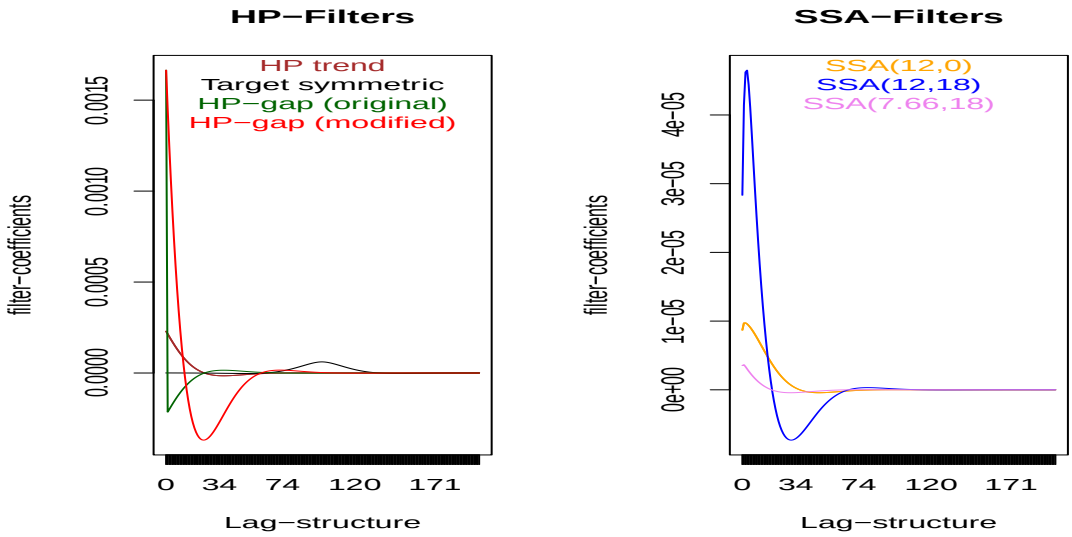


Figure 2.7: HP concurrent (left panel) and SSA concurrent filters (right panel). All filters are arbitrarily scaled to unit-variance.

The coefficients of the specified concurrent and target filters are displayed in fig.2.7: all filters have length $L = 200$ and are normalized to unit-variance ($\frac{1}{L}\sum_{k=0}^{L-1}(b_k - \bar{b})^2 = 1$). The characteristic tips of the SSA-filters are indicative of the boundary constraint $b_{-1} = 0$, see theorem 1. The first two (orange, blue) have identical holding-times (smoothness) but we expect different lead-lag properties (timeliness); the third (violet) has a shorter holding-time matching HP-trend; also, SSA(12,0) is virtually indistinguishable from the MSE-estimate of the symmetric HP-target, not shown here, so that it may be considered as a third benchmark in our analysis. Fig.2.8 compares amplitudes and phase-lags[11] of all concurrent filters, except the original HP-gap which is discarded from further consideration when working with differenced data.

[10] $b_k^{gap}$ is a bandpass design with the property that $\sum b_k^{gap} = 0$ which follows from the definition 2.45. Therefore $b_k^{\Delta gap} = \sum_{i=1}^{k} b_i^{gap} \to 0$ for increasing $k$.

[11] The phase-lag at a given frequency $\omega$ measures the shift, in time-units, between output and input of the filter when fed with a sinusoidal of that frequency.

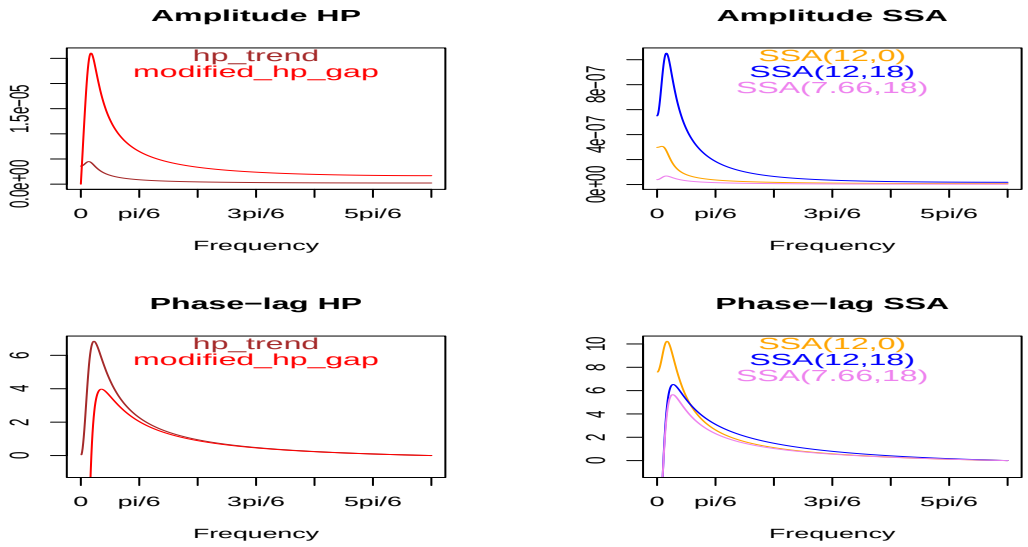


Figure 2.8: Amplitude and phase-lag functions of HP-gap (red), HP-trend (brown), SSA(12,0) (orange), SSA(12,18) (blue) and SSA(7.66,18) violet)

The amplitude functions suggest that all filters, except HP-gap, are lowpass designs; $b_k^{\Delta gap}$ is a bandpass which suppresses low-frequency content: this specific characteristic of HP-gap can be related to the phenomenon of so-called 'spurious cycles', see for example fig.2.12, left panels. Amplitude functions of SSA with larger holding-times (orange and blue) are smallest at higher frequencies, due to stronger smoothing. The phase-lag of HP-gap is smallest overall; SSA(7.66,18) (orange) outperforms HP-trend uniformly at business-cycle frequencies; SSA(12,18) outperforms HP-trend only at lower cycle-frequencies, see fig.2.9.

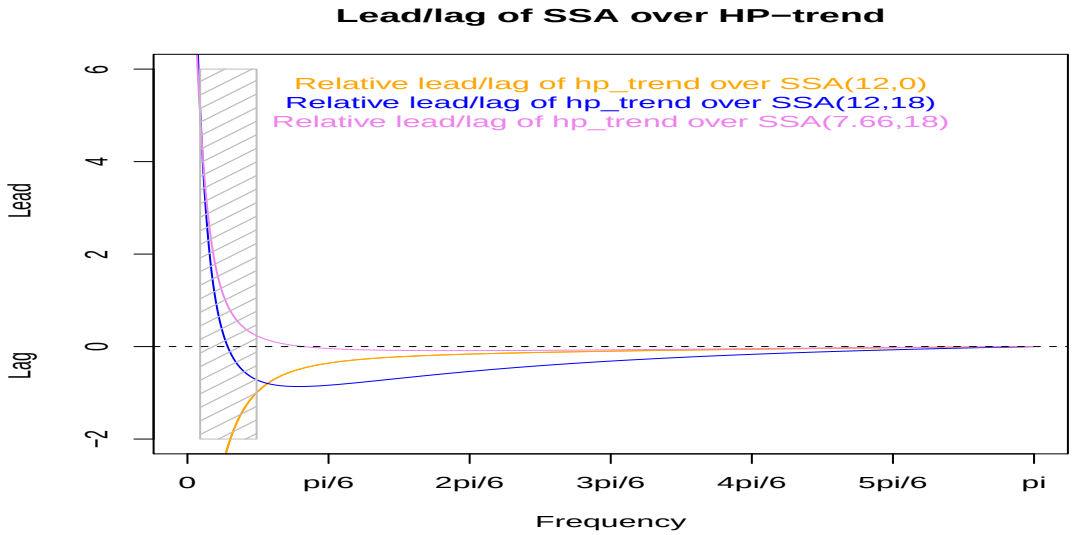


Figure 2.9: Difference of phase-lags of SSA vs. HP-trend. Positive values signify a lead of SSA at the corresponding frequency. Business-cycle frequencies i.e. periodicities between two and ten years are highlighted in the shaded area.

An application of SSA and HP-trend to simulated Gaussian noise leads to mean-shift or $\tau$-statistics summarized in table 2.4 (the mean-shift or $\tau$-statistic is discussed in the appendix: it measures the shift of two competing filter-outputs at zero-crossings and a positive mean-shift implies a corresponding lead or left-shift of the reference-filter, averaged over all crossings). In

| | SSA(12,0) | SSA(12,18) | SSA(7.66,18) |
|---|---|---|---|
| Mean-shift Tau-statistic | -1.0 | -0.0 | 1.0 |

Table 2.4: Mean-shift (tau-statistic) of HP-trend as referenced against SSA based on an application to Gaussian white noise: positive numbers suggest a lead or left-shift by the corresponding SSA-design

summary, HP-gap is expected to lead systematically all lowpass designs, but the eventuality of spurious cycles and zero-crossings could affect the analysis; SSA emphasizes three different research priorities whereby HP-trend is at least equaled in terms of holding-time. We now verify the established diagnostics based on empirical data and effective filter-outputs. A detailed analysis of the US business-cycle is proposed in section 2.6.2 and summary-statistics for a selection of additional countries are provided in section 2.6.3.

### 2.6.2 Application to the US-Industrial Production Index

We consider an application of the proposed concurrent filters to first differences $\Delta I_t$ of the (log-transformed) monthly US industrial production index $I_t$ plotted in fig.2.10: the series effectively starts in 1919-02-01 (FRED database) but we display shorter samples for ease of visual inspection.

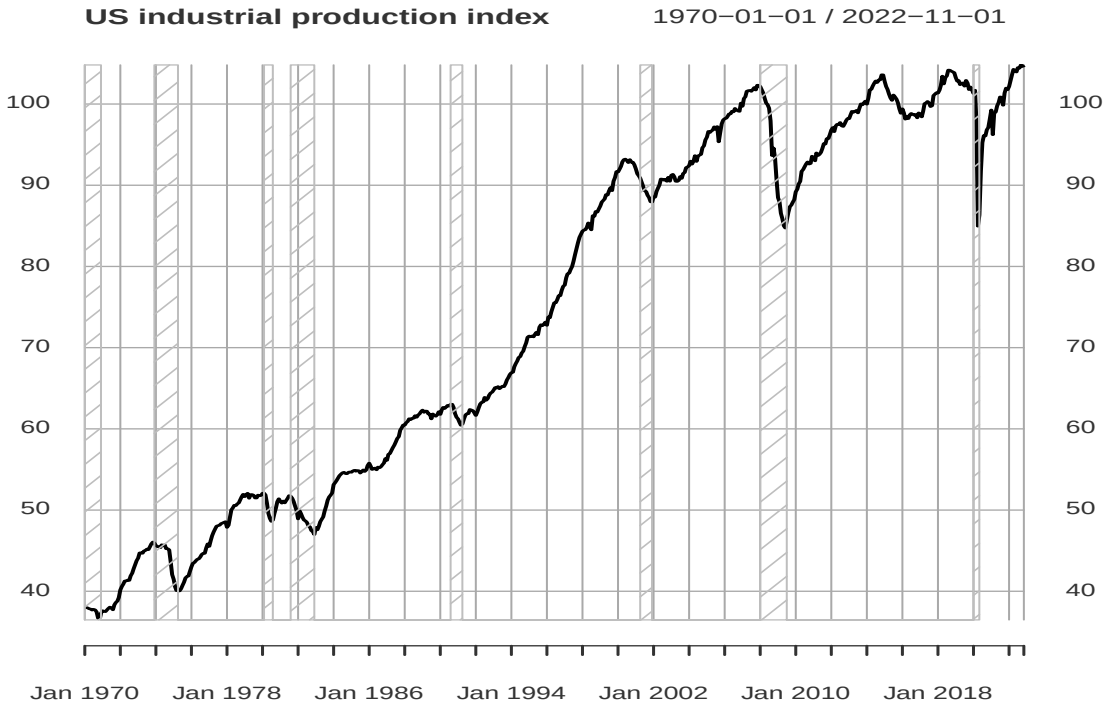


Figure 2.10: Monthly US industrial production index (INDPRO) and recession episodes as dated by the National Bureau of Economic Research, NBER (shaded)

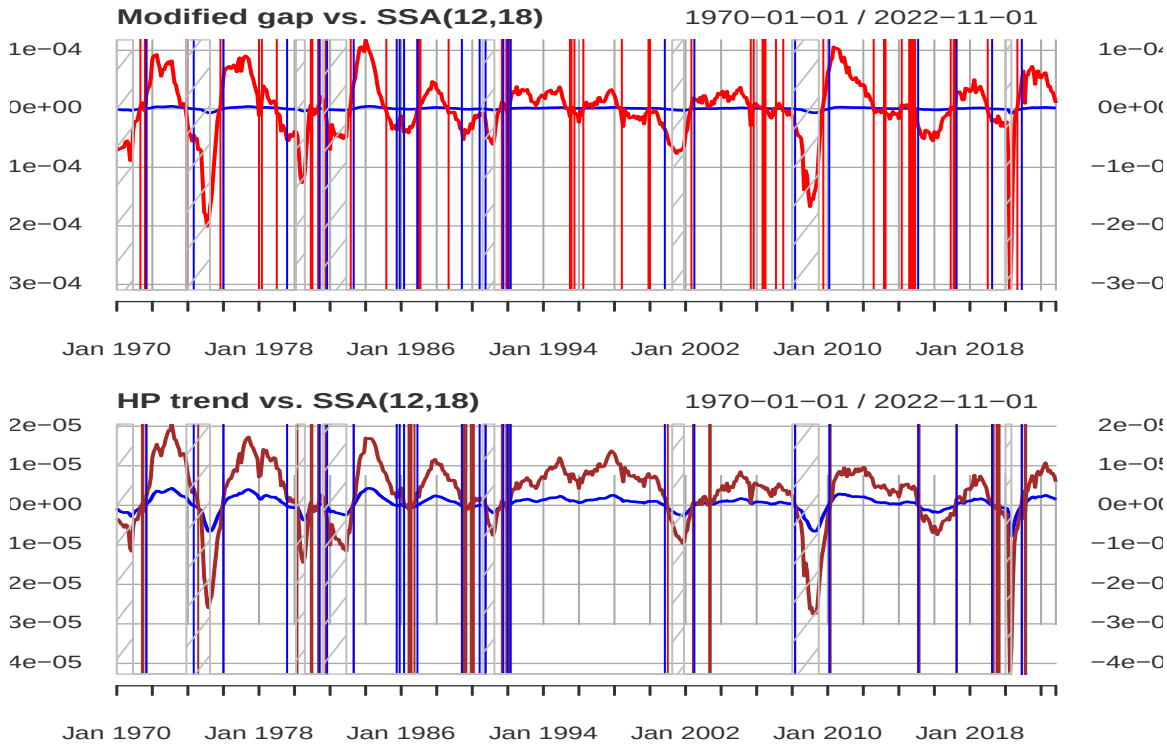


Figure 2.11: Filter-outputs: HP-gap (red) and SSA(12,18) (blue) in the top panel; HP-trend (brown) and SSA(12,18) (blue) in the bottom panel with shaded recession episodes.

Filter-outputs, scaled to unit-variance, are displayed in fig.2.11. Zero-crossings of the filters are marked by vertical lines with matching colors (overlaps are possible). The figure illustrates that the bandpass HP-gap (top panel) generates excessively many crossings, see table 2.5, as well as false systematic sign changes, i.e. spurious cycles, during longer up-swings covering the great moderation, from the early nineties up to the financial crisis, see also fig.2.12, left panels. In contrast, the lowpass designs track the longer cycle-dynamics better. All filters indicate a slowdown of industrial production in 2015 and 2016, at a time when the price for crude oil declined sharply, hence affecting petrol extraction as well as collateral industrial activity in the US. A potential advantage of imposing stronger smoothness can be seen in fig.2.12, right panels, which highlight the pandemic 'great-lockdown' crisis: zero-crossings of the SSA-design are fewer and dynamics are less noisy than the benchmark, which facilitates a real-time assessment of economic conditions, at least in the industrial sector. The bottom-right panel illustrates the potential right-shift or lag of SSA(12,0) at zero-crossings, as measured by $\tau$ in the appendix. Table 2.5 compares timeliness and smoothness performances in terms of $\tau$ and number of sign changes: HP-gap and SSA(7.66,18) outperform in terms of timeliness followed by SSA(12,18), HP-trend and SSA(12,0); but the latter SSA(12,0) outperforms in terms of smoothness, followed by SSA(12,18), SSA(7.66,18), HP-trend and finally HP-gap. These rankings mostly conform with the diagnostics established in the previous section, based on amplitude and phase-lag functions. Based on theoretical as well as empirical

evidences, we now discard HP-gap, subject to spurious cycles, as well as SSA(12,0), subject to a relative lag (recall that SSA(12,0) is virtually indistinguishable from MSE here).

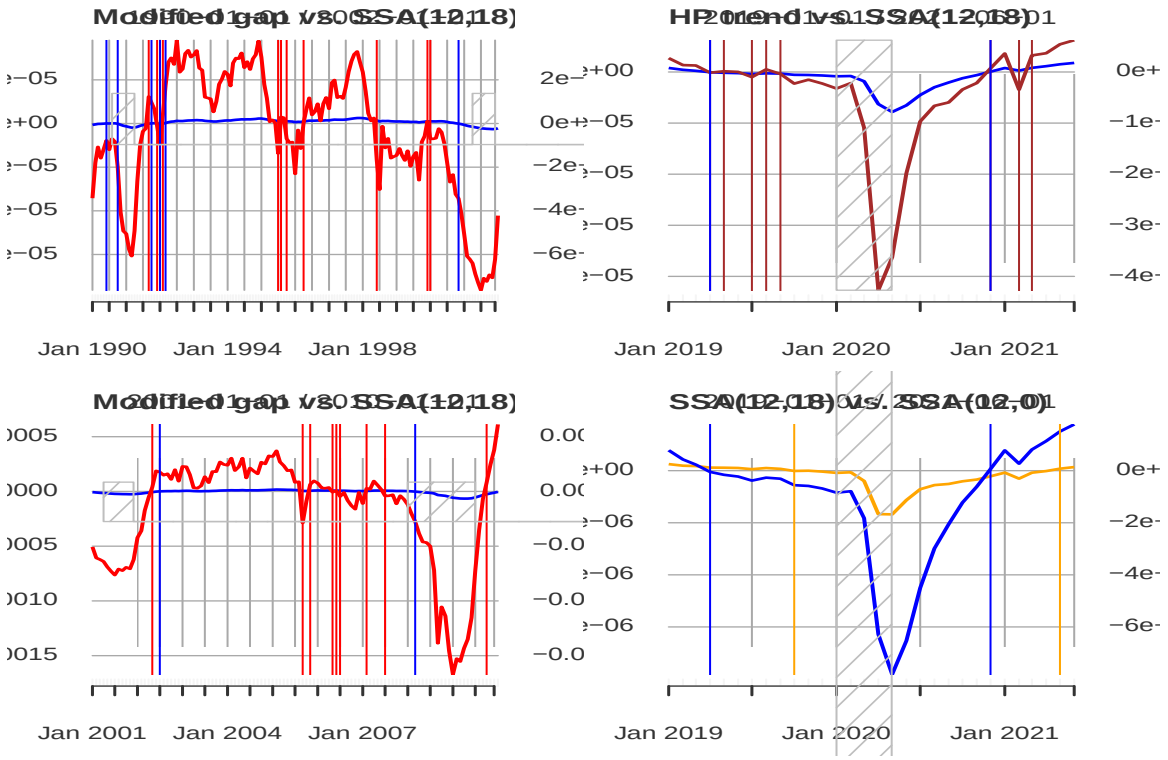


Figure 2.12: Filter-outputs: Modified HP-gap (red), HP-trend (brown), SSA(12,18) (blue) and SSA(12,0) (orange). From 1990-2002 (top left panel, great moderation), 2000-2010 (bottom left panel) and 2019-2021 (right panel: great lockdown). Zero-crossings are marked by corresponding vertical lines (overlaps are possible).

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

Table 2.5: Mean-shift (tau-statistic), as referenced against SSA(12,18), and number of crossings for US-INDPRO from 1935-09-01 to 2022-11-01: positive mean-shifts signify a left-shift or lead by SSA(12,18)

To conclude our analysis of the US business-cycle, we note that the empirical holding-time of 24 months of SSA(12,18) exceeds $ht_1 = 12$. This discrepancy is mainly due to positive autocorrelation (the 'cyclical' growth-rates of the series are smoother than white noise) and the observed effect could be addressed by the extension in section 2.2.2, fixing the link between expected and empirical holding-times and reestablishing interpretability of the hyper-parameter in its original meaning

(but we excluded data-fitting at the outset). Also, the NBER declared 14 recessions in a time span from 1935-09-01[12] to 2022-11-01 which corresponds to a mean-length of expansion-recession cycles of 79 months or roughly 7 years. In comparison, twice the empirical holding-time of the SSA-filter corresponds to $2 \cdot 24$=48 months or four years. Therefore, additional fine-tuning of the design, including the selection of the (HP) target or of the economic indicator[13] or of $ht_1$ in the SSA-specification might be envisioned to match cycle-lengths, at least when calibrated against NBER recession datings.

### 2.6.3 Multi-National Perspective

We here extend the above framework 'as is' to a selection of countries, restricting attention to 'mature' and large economies with correspondingly long and stable cycle histories, ideally differing from the US. see table 2.6. Aggregate performances, obtained by concatenating all series, are summarized in table 2.7: they confirm expectations, as entailed by the selected hyperparameters, and the two SSA-designs are representative of two particular non-exhaustive research priorities. While the lag of HP-trend against the reference SSA(12,18) is statistically insignificant, the latter outperforms the former in terms of smoothness. The lead of SSA(7.66,18) over the reference SSA(12,18), and thus over the benchmark, is strongly significant: for similar smoothness, SSA outperforms HP-trend in terms of timeliness (the strict equality of the total number of crossings is fortuitous in this case). We may also refer to fig.2.13 in the appendix for a visualization of $\tau$ and the underlying statistical t-test. While our treatment of timeliness, by alterations of the forecast horizon $\delta$, might be felt as 'ad hoc', we argue that the intended effect can be established, measured and tested for statistical significance; moreover, the upshot of our proceeding consists in interpreting the forecast horizon as an additional free tuning-parameter, affecting properties of the predictor to match alternative research priorities. In this abstract perspective, a more legitimate formal approach can be developed for addressing prediction-tradeoffs in a fundamental manner and SSA is a first step pointing to this direction.

## 2.7 Conclusion

We propose a novel SSA-criterion which emphasizes sign accuracy and zero-crossings of the predictor subject to a holding-time constraint. Under the Gaussian assumption, the classic MSE-criterion is equivalent to unconstrained SSA-optimization: in the absence of a holding-time constraint and down to an arbitrary scaling nuisance. We argue and demonstrate that the proposed concept is resilient against various departures of the Gaussian assumption. Moreover, the approach is interpretable and appealing due to its actual simplicity and because the criterion merges relevant concepts of prediction in terms of sign accuracy, MSE, and smoothing requirements. The business-cycle case illustrates application to an existing benchmark in view of modifying predictors according to particular research priorities. In this context, timeliness and smoothness of real-time

[12] The earliest observations are skipped due to filter initialization.

[13] Some of the downturns of the industrial production index, such as in 2015 and 2016, do not classify as economic recessions and therefore a direct comparison of corresponding 'cycles' is subject to caution.

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

Table 2.6: Number of crossings and mean-shift (tau-statistic), as referenced against SSA(12,18): a positive shift means a corresponding lead.

designs can be controlled in an effective and interpretable way.

## 2.8 Appendix

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

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

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

$$\tau(N,n) := \frac{1}{|ZC_N|} \left( \sum_{j=1}^{|ZC_N^+|} (t^+_{f(j,N),n} - t^+_{jN}) + \sum_{j=1}^{|ZC_N^-|} (t^-_{f(j,N),n} - t^-_{jN}) \right) \tag{2.46}$$

where $f(j,N)$ is the index of the zero-crossing of $y_{tn}$ closest to $t^+_{jN}$ or $t^-_{jN}$ in the corresponding subsets $ZC_n^+$ or $ZC_n^-$ i.e. $|t^+_{f(j,N),n} - t^+_{jN}| = \min_i |t^+_{in} - t^+_{jN}|$ and similarly for the downturns. Note

| | SSA(12,18) | HP trend | SSA(7.66,18) |
|---|---|---|---|
| Mean-shift over countries | 0.00 | 0.14 | -0.43 |
| t-test for time-shift | 0.00 | 1.25 | -9.54 |
| Total number of crossings | 304.00 | 494.00 | 494.00 |

Table 2.7: Aggregate mean timeliness and smoothness performances obtained by concatenation into a single long series. Shifts at zero-crossings i.e. tau-statistics are referenced against SSA(12,18): positive shifts indicate a lead or left-shift of the reference; t-statistics for significance of the lead or lag are reported in the middle row

that sums are taken over zero-crossings of the reference filter $y_{tN}$, which is always a SSA-design in the empirical sections. We recommend that $y_{tn}$, $n = 1, ..., N$ should have 'similar' crossings for the comparison to be meaningful (comparing a lowpass to a highpass would lead to difficulties when interpreting results); also, ideally, the reference filter should be smoother, with fewer zero-crossings, as is the case in our examples. In our BCA application, we mainly rely on SSA(12,18): we prefer this reference to the proper target, i.e. the HP-symmetric filter, because the latter cannot be used towards the sample-end, thus excluding the latest and important great-lockdown crisis, see fig.2.12. Also, timeliness performances of all concurrent filters can be assessed against another causal filter, which facilitates direct comparisons.

The $\tau$-statistic 2.46 is called *mean shift* of the reference filter $N$ with respect to filter $n$; the former is called *leading* or *lagging*, with respect to the latter, depending on the mean-shift being positive or negative. The statistic could be split into separate downturn and upturn sums in the case of asymmetry. A classic t-test can be used to infer statistical significance of a mean-lead or a -lag at zero-crossings, assuming the summands in 2.46 to be independently distributed. For illustration, fig.2.13 displays the cumulated shifts at zero-crossings of the filters in table 2.7 (single long concatenation of all country-specific series).

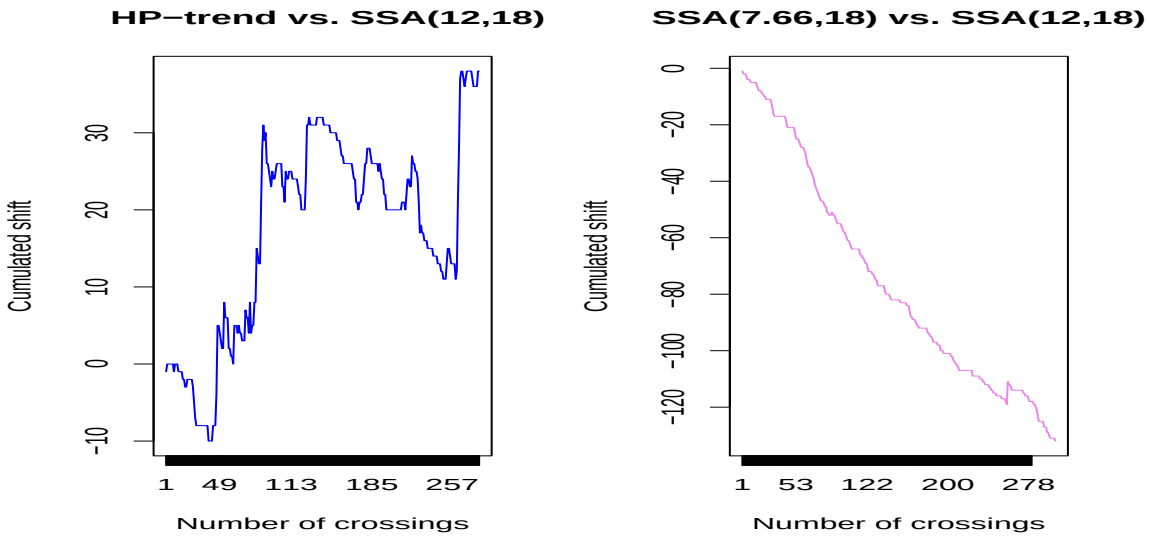


Figure 2.13: Cumulated Shift at zero-crossings computed on concatenated series: HP-trend (left panel) and SSA(7.66,18) (right panel) are both referenced against SSA(12,18). An upward trend signifies a lead of the reference SSA(12,18).

The mean-shift or $\tau$-statistic corresponds to the slope of these curves and a positive slope signifies a lead of the reference design. The t-statistic tests the alternative (non-vanishing drift) against the null (vanishing drift): the large absolute t-value for the second SSA-filter in table 2.7 reflects the strong drift in the right panel of figure 2.13. Finally, note that the time-shift statistic considers differences of *closest* crossings only, i.e. earlier or later 'noisy' sign-changes of the contender $y_{tn}$ are ignored. Therefore, $\tau$ tends to be biased against the reference-filter $y_{tN}$ if the latter is smooth and leading, as is mostly the case of SSA-designs in our application. As an example, the $\tau$-statistic of HP-trend referenced against SSA(12,18) in the top-right panel of fig.2.12 would vanish over the latest 'great-lockdown' crisis, because the closest crossings of HP-trend overlap with SSA. But the figure also indicates the presence of a couple of delayed 'noisy' sign-changes by HP, to the right of the reference SSA-crossings, which are ignored by our statistic. In practice, analysts would typically wait for a confirmation of a sign-change, up to a point where a relapse can be excluded with some confidence. Since these considerations are wholly ignored by $\tau$, the mean-shift can be considered as a conservative measure for the lead of SSA in our examples.

## 2.9 Conclusion

We propose a novel SSA-criterion which emphasizes sign accuracy and zero-crossings of the predictor subject to a holding-time constraint. Under the Gaussian assumption, the classic MSE-criterion is equivalent to unconstrained SSA-optimization: in the absence of a holding-time constraint and down to an arbitrary scaling nuisance. We argue that the proposed concept is resilient against various departures of the Gaussian assumption. Moreover, the proposed approach is interpretable

and appealing due to its actual simplicity and because the criterion merges relevant concepts of prediction in terms of sign accuracy, MSE, and smoothing requirements. Alternative research priorities, such as timeliness and smoothness, can be addressed consistently and effectively by a pair of hyper-parameters and the smoothness or holding-time constraint has a natural and interpretable meaning.

# Bibliography


[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] Granger, C.W.J. (1966). The typical spectral shape of an economic variable. *Econometrica* **34**, 150-161.

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

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

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

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

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

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

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

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

[14] Osterrieder, J. (2017) The Statistics of Bitcoin and Cryptocurrencies. *Advances in Economics, Business and Management Research (AEBMR)* **Vol. 26**.

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

[16] Wildi, M. (2023) ???