Smooth and Persistent Forecasts of German GDP: Balancing Accuracy and Stability*

Katja Heinisch† Simon van Norden‡ Marc Wildi§

August 2025

Work in progress, not for circulation

Abstract

This paper explores the trade-off between informational efficiency and forecast smoothness in economic forecasting. Forecasts that minimize mean squared forecast errors often exhibit excessive volatility, limiting their practical applicability. We propose the Multivariate Smooth Sign Accuracy (M-SSA) framework as a solution, extracting smoothed components from leading indicators. This enhances the signal-to-noise ratio and controls forecast volatility and timing. Applied to German GDP growth, our method yields smoothed forecasts that improve the forecasting accuracy of traditional models, particularly over medium-term horizons. While smoother forecasts tend to lag slightly, this drawback can be mitigated by adjusting the forecast horizon. These findings highlight the practicality of the M-SSA framework for both forecasters and policymakers, especially in settings where updating forecasts or policy adjustments are costly.

Keywords: Forecast Smoothing, Smooth Sign Accuracy, Time Series Filtering
JEL Classification: C53, E37, E66

*We would like Barbara Rossi and the participants at the Vienna Forecasting Workshop 2025 for their comments and suggestions.

†Halle Institute for Economic Research (IWH).
‡HEC Montréal and CIREQ.
§Zurich University of Applied Sciences (ZHAW).

1 Motivation

> "As we parse the incoming information, we are focused on separating the signal from the noise as the outlook evolves."
> *Speech by Federal Reserve Board Chair Jerome H. Powell, March 07, 2025, University of Chicago Booth School of Business*

Economic forecasters face a dilemma: while informational efficiency requires that forecasts incorporate all available data, doing so often results in forecasts that are excessively volatile. In addition, forecasts may reflect excessive "noise" due to overfitting. Further, forecast users (including public and private decision makers) may face constraints in their speed of adjustment to changing forecasts. For example, fixed annual budgets may constrain spending, or transaction costs may limit changes in investments. Others may attempt to smooth their reactions for reasons of signalling. Important examples include dividend smoothing by corporate managers and interest rate smoothing by central banks.

Wildi (2024, 2025) and McElroy and Wildi (2019, 2020) propose novel methods that optimally smooth forecasts by controlling the expected frequency of sign changes. The resulting Smooth Sign Accuracy (SSA) framework encompasses efficiently the mean-square error (MSE) as a special case and produces increasingly smooth forecasts as changes in the forecast direction (sign) are increasingly penalized.

This study extends this framework by replacing raw leading indicators with smoothed and filtered components, thereby improving the signal-to-noise ratio in the predictor variables. We apply this framework to well-established German economic indicators to show its ability to improve forecasts of GDP growth (Lehmann and Reif, 2021; Heinisch and Scheufele, 2018) and GDP trends using a variety of sample periods and performance metrics.

We find that smoothing indicator-based forecasts often provides useful leading information for GDP, especially over medium-term horizons ranging from two to four quarters ahead. We also show that while smoother GDP forecasts lag efficient MSE forecasts slightly, this can be mitigated by adjustments in the forecast horizon.[1]

We proceed as follows. Section 2 describes our data series, the sample period used, and documents their dynamic behaviour. Section 3 provides benchmark forecasts using simple direct projections of GDP based on selected indicators. We demonstrate how in-sample forecast performance may improve when we "de-noise" the indicators prior to projection with simple smoothing filters. Section 4 defines the M-SSA method and shows how its forecast performance compares to that of direct projections as well as how it varies with forecast horizon. Section 5 considers the ability of our indicators to forecast German GDP growth out-of-sample using the M-SSA framework, and Section 6 investigates the temporal stability of the estimated forecasting equations. Section 7 summarizes our results and concludes.

[1] The lag reflects the trilemma highlighted by McElroy and Wildi (2019) in which forecasters choose between forecast accuracy, timeliness, and volatility.

2 Data and Dependence

2.1 Data and Publication Lags

The analysis employs a set of well-established leading indicators to forecast quarterly German GDP (Drechsel and Scheufele, 2012; Heinisch and Scheufele, 2018), including industrial production (IP), the ifo Business Climate Index (ifo_c), the Economic Sentiment Indicator (ESI), and the term spread between 10-year bond yields and the 3-month EURIBOR rate (spre_10y_3m) covering the period 1995 to 2024. We take into account the publication lag of the indicators and the GDP series (Table 1).[2] While GDP is available with a one-quarter delay, and IP is released with a two-month lag, business surveys and financial data are published without any lag.

Table 1: Ragged edge structure. As of January 2025, before publication of the GDP flash estimate.

	GDP	IP	ifo_c	ESI	spr_10y_3m
2024-07-01		90.700	85.800	92.500	-1.300
2024-08-01		93.200	84.900	90.900	-1.400
2024-09-01	902.571	91.200	83.700	89.800	-1.400
2024-10-01		90.800	84.100	90.700	-1.200
2024-11-01		92.200	83.800	89.300	-0.900
2024-12-01			82.800	86.900	-0.800
2025-01-01			82.500	88.100	-0.200

To allow for missing observations due to publication lags, we re-align series to reflect the information available to forecasters. Table 2 displays the artificial result as of the end of 2024, with GDP and IP columns shifted to reflect their release dates rather than the periods that they measure. This data configuration will be the used for nowcasting or forecasting GDP. For illustrative purposes, an additional target column is included on the right side of the table, representing the dependent variable for nowcasting applications.[3]

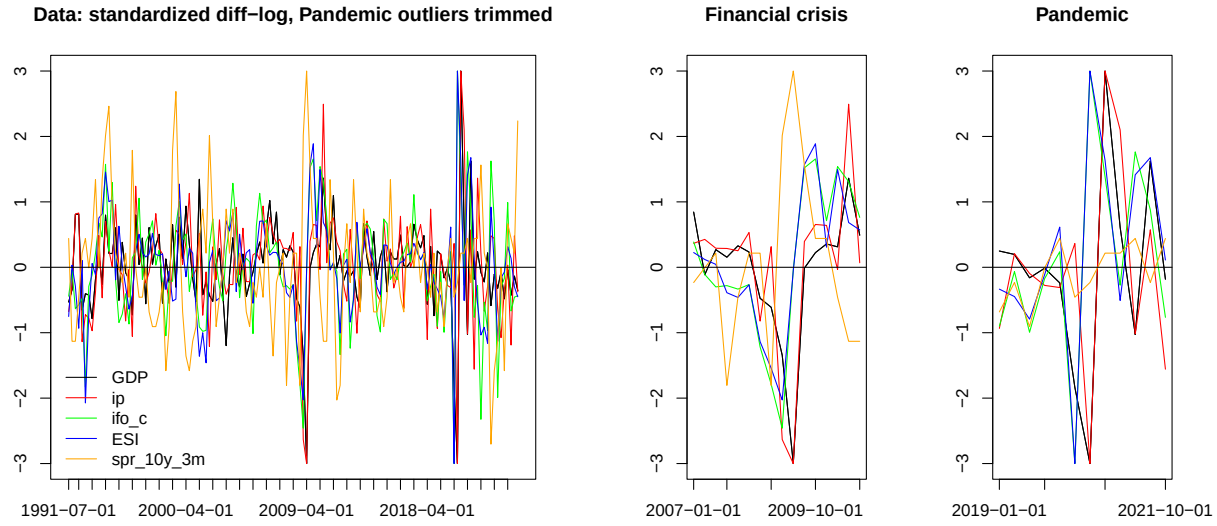
Table 2: Aligned dataset. At each date, various columns show the most recent values of each predictor available to forecasters. Column 6 represents the target variable for a GDP nowcast (i.e. the values of column 1 shifted upwards by the publication lag of GDP.

	GDP	IP	ifo_c	ESI	spr_10y_3m		Target (nowcast)
2024-07-01		93.400	85.800	92.500	-1.300	—	
2024-08-01		90.700	84.900	90.900	-1.400	—	
2024-09-01	901.623	93.200	83.700	89.800	-1.400	—	902.571
2024-10-01		91.200	84.100	90.700	-1.200	—	
2024-11-01		90.800	83.800	89.300	-0.900	—	
2024-12-01	902.571	92.200	82.800	86.900	-0.800	—	?

To ensure comparability and mitigate scale effects, all indicators (except the term spread, which is differenced only) are log-differenced and standardized. Additionally, extreme outliers associated with the COVID-19 pandemic are trimmed to prevent distortions in subsequent figures. Monthly series are shifted (Table 2) and averaged to quarterly series (Figure 1a), while Figure 1b offers

[2] Data revisions are not considered, given that previous studies (Heinisch and Scheufele, 2019) have shown that their overall effect on forecast performance is minor.

[3] In the case of an h-step ahead forecast, this column is shifted upwards by h quarters.



(a) Full Sample.

(b) Leads & lags during financial crisis & Pandemic.

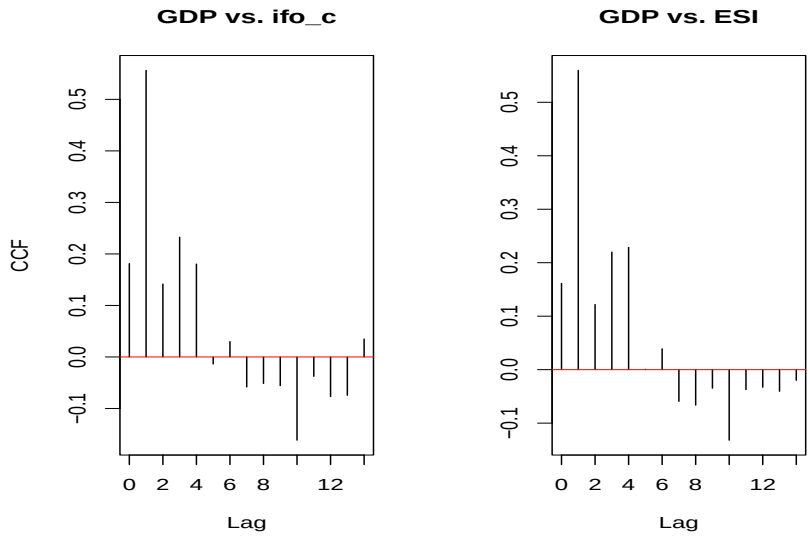
Figure 1: GDP & Indicators: standardized log-differences, trimmed to ±3 standard deviations.

a close-up view of the financial crisis and the pandemic. The latter figure shows that the real-time GDP and industrial production series are synchronized at turning points, whereas the other indicators tend to lead somewhat. The challenge for forecasters is to use the latter in a multivariate approach to improve forecasting performance over that of univariate benchmarks.

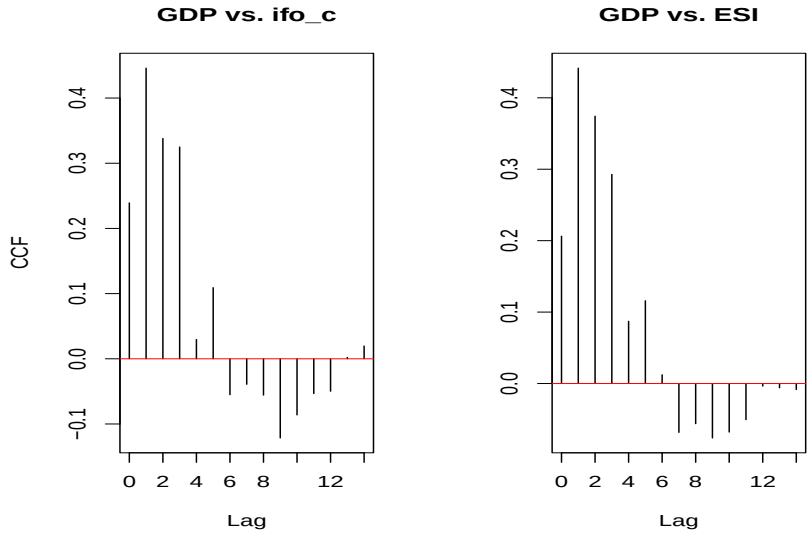
2.2 Dependence

To examine the joint dynamics of our variables, we estimate their cross-correlations as well as their vector moving-average (VMA) representations. The latter is a key input to the multivariate filters we develop in Sections 4 and 5.

Figures 22 (full sample) and 23 (omitting the Pandemic period) show the sample cross-correlation functions (CCF's) between current and lagged GDP and the two survey indicators. Both figures show strong correlations that peak when GDP is slightly lagged. Results for IP showed no such lag, while correlations for the term spread were weaker. The figures also show that the precise strengths of the correlations are influenced by the outliers associated with the COVID-19 pandemic. In the interest of model stability, we exclude data from Q4-2019 to Q4-2020 (pandemic) when modelling the dynamics for the multivariate filters developed below.



(a) Full data set from Q2-1991 to Q4-2024.



(b) Without Pandemic (omitting Q4-2019 to Q4-2020).

Figure 2: Sample cross-correlation functions (CCF) showing correlations with lags of GDP growth.

We then estimated a variety of parsimonious VARMA models to capture these dynamics, selecting parsimonious VAR(1) and VAR(3) specifications.[4] The figures 3 and 4 show the implied responses of GDP to shocks to each indicator variable.[5]

[4] We used the `refVARMA()` procedure in the Tsay et al. (2022) `MTS` package, which sets VARMA coefficients with t-statistics below a critical threshold to zero (see Tsay, 2013). This led to a VAR(1) specification with a threshold of 1.5. The VAR(1) left some residual evidence of higher-order dynamics, but capturing these in our limited sample size was difficult without overfitting. We therefore explored VAR(p) models with coefficients estimated by Elastic Net using the Zou and Hastie (2020) `elasticnet` package. This led to a VAR(3) specification with 7 steps. For additional details, see Zou and Hastie (2020) and the references therein.

[5] Shown are the MA coefficients for GDP from the vector moving-average (VMA) models obtained by inversion of the VAR(1) and VAR(3) representations respectively (i.e. the impulse-response functions (IRFs) of GDP to the correlated reduced-form shocks.)

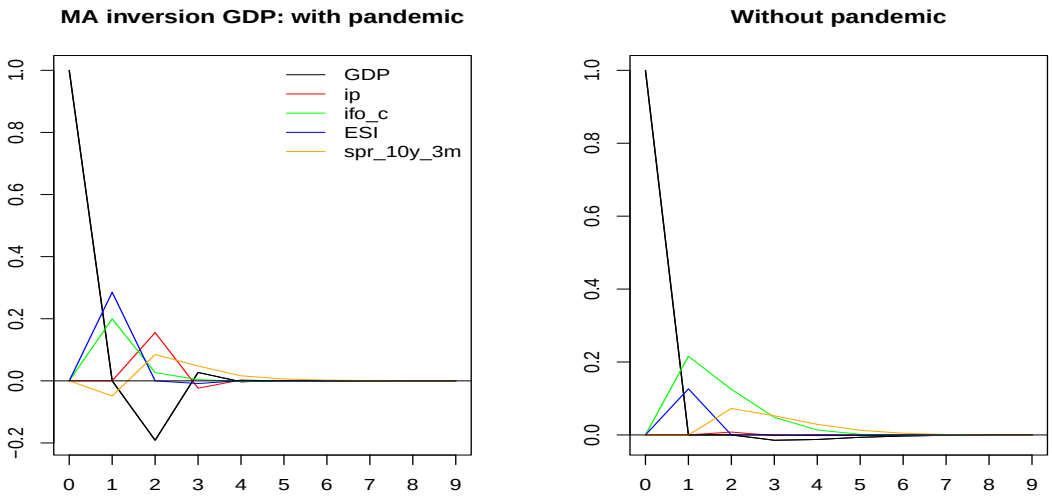


Figure 3: VMA coefficients for GDP implied by VAR(1)
Estimation with pandemic (left) and without pandemic (right).

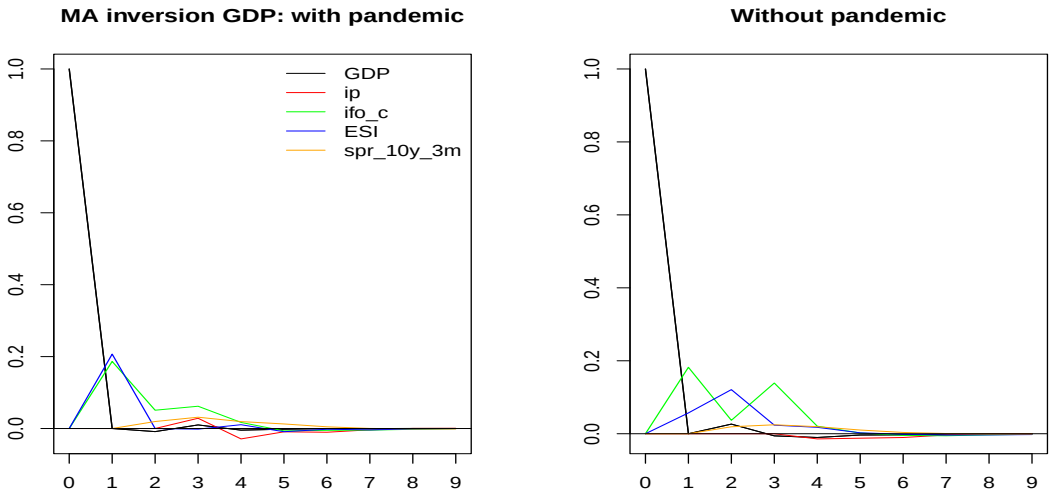


Figure 4: VMA coefficients for GDP implied by VAR(3)
Estimation with pandemic (left) and without pandemic (right).

GDP exhibits the strongest response to shocks in the two survey indicators—ifo and ESI.[6] The impact of the pandemic modifies the lag structure by emphasizing shorter dependencies triggered by the alternating outliers during that specific episode. Removing the pandemic from the analysis results in smoother, somewhat longer-tailed dependencies, consistent with the observed empirical autocorrelation functions. In particular, the VAR(3) model captures slightly longer lead times and more nuanced lag patterns.[7]

[6] There is also weaker evidence of a response to the term spread at longer lags.

[7] The VAR(3) model yields marginal improvements in multivariate filtering performance. However, due to its relative simplicity, this study primarily employs the VAR(1) model. The M-SSA package (accessible at

The delayed response of GDP to movements in the survey indicators suggests that multivariate models could yield better forecasts than those of univariate model, a hypothesis we explore in the next section.

3 Forecasts with Pre-Filtering

In this section, we demonstrate how pre-filtering indicator variables may improve the performance of simple forecasts by removing noise. We then demonstrate how in-sample forecast performance improves when we "de-noise" the indicators prior to projection with simple smoothing filters.[8]

For simplicity, all the models in this section are direct projections of GDP on our indicators and take the form

$$u_{t+h} = \boldsymbol{\beta'} \cdot \mathbf{X}_t + \varepsilon_t \tag{1}$$

where

- u_{t+h} is the quarterly growth rate of GDP for forecast horizon $h = \{0, 1, 2, ...\}$
- $\mathbf{X}_t \equiv [1, u_{t-1}, ip_t, ifo\_c_t, ESI_t, spr\_10y\_3m_t]'$ or their transformed counterparts and u_{t-1} accounts for the publication lag of GDP.[9]
- $\hat{\boldsymbol{\beta}}$ is estimated by least-squares, omitting observations affected by the pandemic, i.e. Q4-2019 to Q4-2020.

`https://github.com/wiaidp/R-package-SSA-Predictor`) offers more advanced capabilities, including the implementation of the VAR(3) model and Bayesian VAR specifications.

[8] The smoothing filters are one-sided (concurrent) filters of length 31, therefore, estimation of both forecasting models omits the first 30 observations.

[9] Including lagged GDP enables the nesting of the AR forecast within the broader modeling framework.

3.1 Unfiltered Indicators

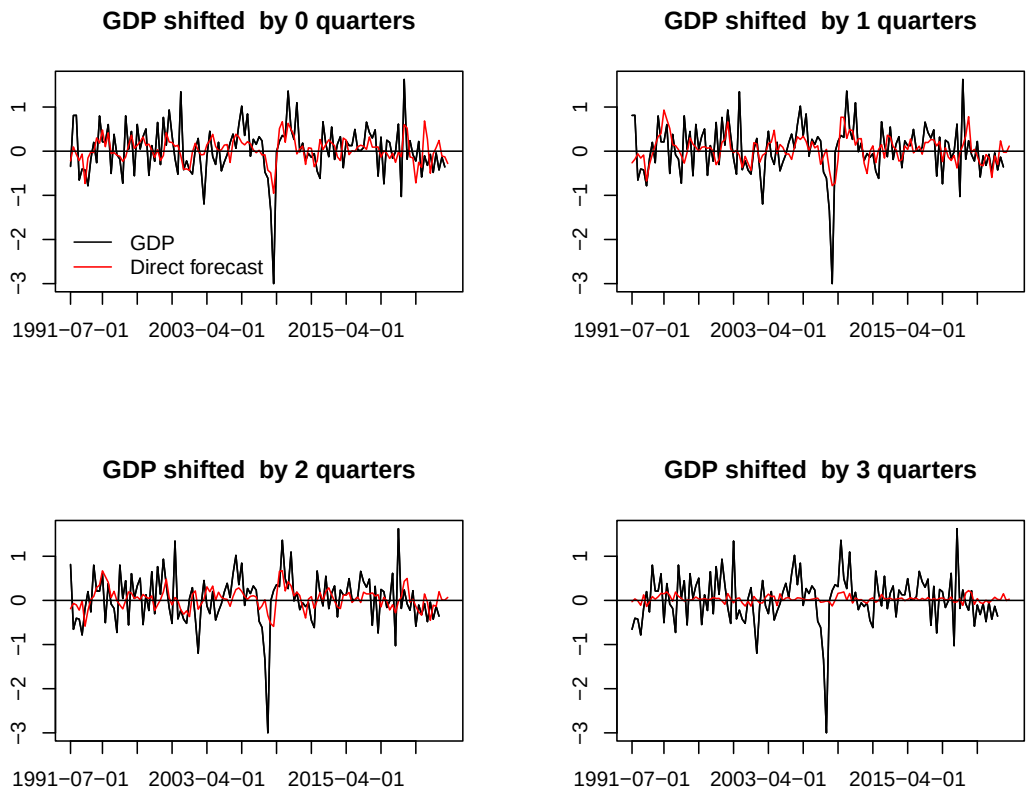


Figure 5: Direct forecasts (red lines) based on all indicators and full sample information (without pandemic.)
Target (black line): GDP left-shifted by 0 - 3 quarters.

Figures 5 and 6 compare actual and fitted values for GDP at forecast horizons $h = 0, \ldots, 3$, using the four indicators and the target, i.e., GDP shifted h quarters ahead. Figure 5 shows results over the sample without the pandemic, while Figure 6 provides a closer look at the fit during the financial crisis. Table 3 shows p-values for tests of the joint null hypothesis that all coefficients on the indicators are zero (HAC-adjusted Wald test).

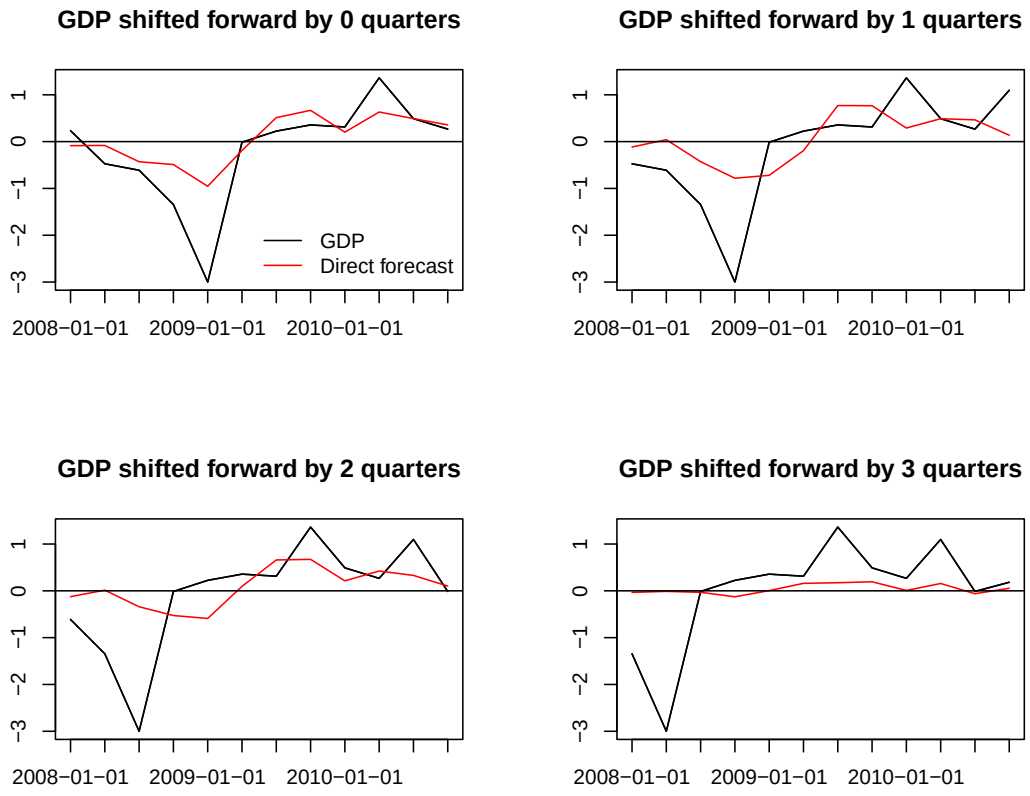


Figure 6: Direct forecast across the financial crisis: regression based on all indicators and full sample information, without pandemic.

The results indicate that as the forecast horizon h increases, forecast become less volatile and their turning points (i.e., peaks and troughs) become increasingly right-shifted (lagged) relative to the target GDP. Additionally, the p-values of the HAC-adjusted Wald test precipitously increase from $h = 2$ to $h = 3$ (see Table (3)), implying that the indicators become statistically insignificant for forecast horizons of more than two quarters.[10]

Table 3: p-values for $H_0 : \boldsymbol{\beta} = 0$ in (1) based on HAC-adjusted Wald test. Estimation over full sample (without pandemic.)

	h= 1	h= 2	h= 3	h= 4	h= 5	h= 6
Unfiltered	0.000	0.001	0.910	0.509	0.091	0.178
HP-C filtered	0.001	0.036	0.046	0.049	0.012	0.022

3.2 Direct Forecasts Based on Filtered Indicators

We conjecture that economic indicators may be contaminated by unpredictable high-frequency noise, thereby obscuring the effective 'signal' and making a direct regression more susceptible to

[10] These findings are confirmed by the relative root mean-square errors comparing direct forecasts against the mean benchmark, see Table 8 in the Appendix.

overfitting. If so, it may be possible to highlight the signal by filtering our indicators to dampen high-frequency noise.

The Hodrick-Prescott (HP) filter is a classic tool used in business cycle analysis: the filter is specified by a single smoothing parameter λ. The value $\lambda = 1600$ is the most common value used for business cycle analysis, but Phillips and Jin (2021) note that this removes relevant information due to excessive smoothing. To address this concern, we instead use $\lambda = 160$ to allow for a more adaptive filter.

Figure 7 presents both classic two-sided and one-sided HP(160) (denoted HP-C for *concurrent*) filters (top panels) along with their associated amplitude functions in the frequency domain (bottom panels). It confirms that the HP-C filter assigns greater weight to yearly components—relevant in a medium-term forecast context— than the classic quarterly HP filter, while also reducing undesirable high-frequency noise. Therefore, while the choice of the HP(160) filter in this example is somewhat arbitrary, it should serve the purpose of reducing high-frequency noise while preserving longer-duration movements.[11]

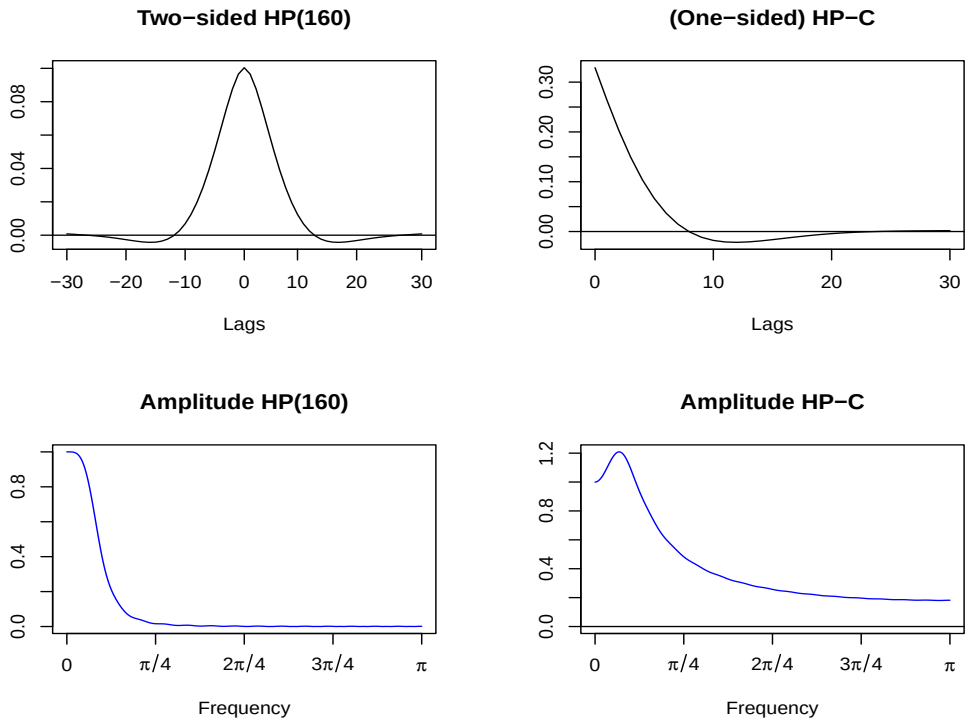


Figure 7: HP($\lambda = 160$) filters
Left: HP (Two-sided, symmetric) Right: HP-C (One-sided, concurrent)
Top Row: Filter Weights Bottom Row: Amplitude Functions.

Figures 8 and 9 provide results comparable to those of Figures 5 and 6 but now using indicators smoothed with the one-sided HP(160), (denoted as the HP-C, for *concurrent*) filter. The fitted values lag GDP slightly less and track turning points somewhat more quickly. Table 3 shows p-

[11] A comprehensive technical analysis of the effect of λ on the resulting GDP predictor can be conducted using the M-SSA package (`https://github.com/wiaidp/R-package-SSA-Predictor`).

values for tests of the joint null hypothesis that all coefficients $\boldsymbol{\beta} = 0$. The use of filtered indicators improves their statistical significance at forecast horizons $h > 2$ quarters.

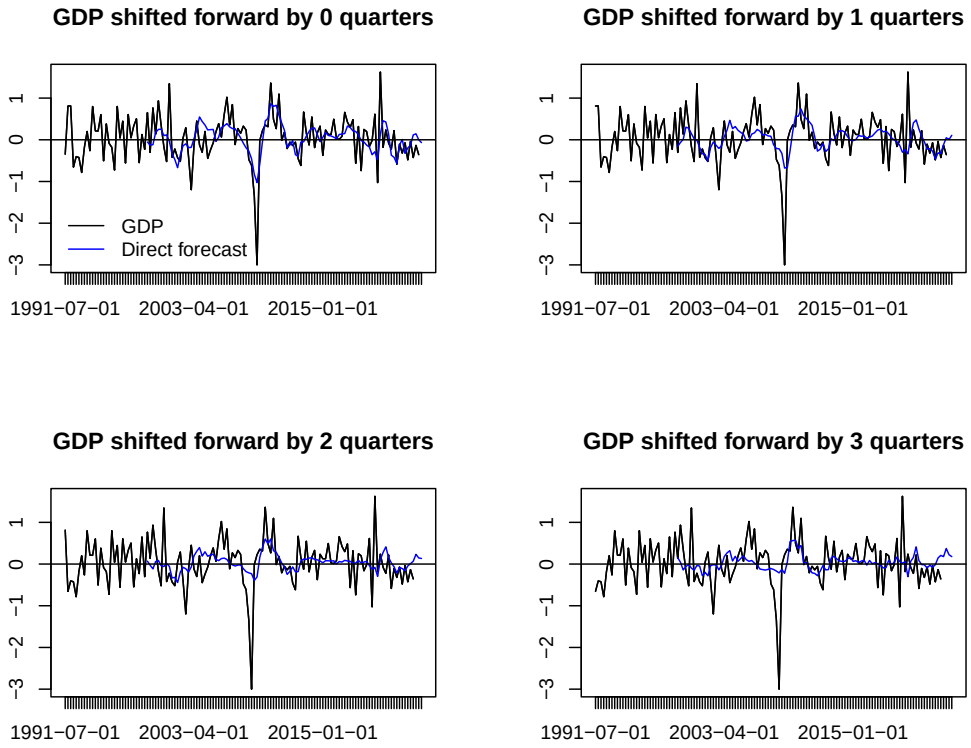


Figure 8: Direct HP forecasts: entire data set with exclusion of the pandemic.

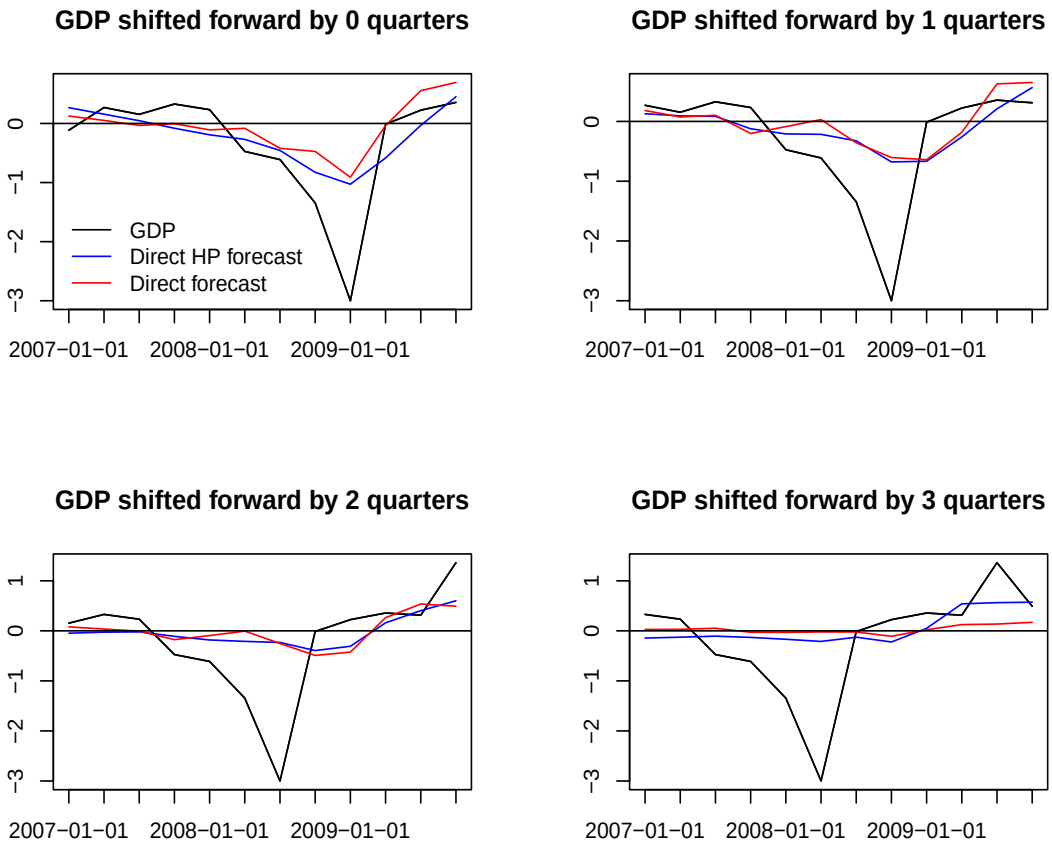


Figure 9: Shifted GDP (black), direct HP forecast (red) and classic direct forecast (blue) over the course of the financial crisis: regression based on all indicators and full sample information (without pandemic).

The application of the one-sided HP-C filter to the indicators yields some measurable improvements in forecast performance. However, the noise suppression capability of the one-sided filter is compromised, as explained by Wildi (2025) and illustrated by the high-frequency leakage of the amplitude function in Figure (7). To address these limitations, we turn to the Multivariate Smooth Sign Accuracy (M-SSA) framework proposed by Wildi (2025).

4 Multivariate Causal Filter: M-SSA

To understand the smoothing properties of the M-SSA framework, we begin with the Smooth Sign Accuracy (SSA) case, which considers a univariate smoothing problem, before turning to M-SSA, its multivariate generalization.[12] In the remainder of the section, we demonstrate the properties of M-SSA nowcasts and forecasts for smoothed GDP.

[12] See Wildi (2024) for further details on the SSA framework and its applications.

4.1 SSA Optimization Criterion

Let $\mathbf{x}_t$ be a time series representing the data, with observations $(x_t, ..., x_1)'$. Let $\mathbf{z}_t = (z_t, ..., z_1)'$ denote a target series, which depends on $\mathbf{x}_t$ as well as on unknown future x_{t+k}, $k > 0$ (and possibly unknown past x_k, $k \leq 0$). For simplicity, we assume stationarity of all time series involved. In our current framwork, z_t corresponds to the output of a two-sided Hodrick–Prescott filter with smoothing parameter $\lambda = 160$ applied to the doubly infinite sequence x_k, $k \in \mathbb{Z}$. The finite sample $\mathbf{x}_t$ is the portion effectively observed and represents the log-differenced GDP (the dependence on $x_0, x_{-1}, ...$ is neglected because the HP filter weights decay rapidly to zero; see Fig. 7). The prediction task comprises tracking z_t using the (causal) predictor $y_t = \sum_{k=0}^{L-1} b_k x_{t-k} = \mathbf{b}'\mathbf{x}_t$, where the filter $\mathbf{b}$ with weights $b_0, ..., b'_{L-1}$ (actually of length $L = t$) is available for optimization purposes.

The task of tracking a target 'optimally' can be formalized in different ways; here, we focus on the correlation $\rho(z, y, h)$ between y_t and target z_{t+h}, where $h \geq 0$ denotes the forecast horizon (backcasting with $h < 0$ is ignored in this context). For simplicity, assume a fixed horizon, say $h = h_0$, so that we may drop the reference to h in our notation. In the univariate case, Wildi (2025) proposes the Smooth Sign Accuracy (SSA) as an optimization criterion for determining $\mathbf{b}$:

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

where $\rho(y, 1)$ is the first-order autocorrelation of the predictor y_t. The parameter ρ_1 controls the smoothness of the predictor: higher values of ρ_1 favour smoother trajectories of y_t. Assuming that the process x_t has mean zero, Wildi (2024) establishes a connection between $\rho(y, 1)$ and the expected duration between consecutive 'zero-crossings' (sign changes) of y_t, the 'holding time', denoted by $HT(y)$[13]:

$$HT(y) = \frac{\pi}{\arccos(\rho(y, 1))}. \tag{3}$$

Since eq. (3) is strictly monotonic, the SSA criterion can be reformulated as the optimization problem:

$$\left.\begin{array}{c} \max_{\mathbf{b}} \ \rho(z, y) \\ \frac{\arccos(\rho(y,1))}{\pi} = 1/HT_1 \end{array}\right\}, \tag{4}$$

where the smoothness parameter $1/HT_1$ expresses the expected rate of sign changes of the predictor.

One implication of criterion (4) is that the objective function as well as the constraint are indifferent to an affine transformation of the predictor. This ambiguity can be resolved by assuming an arbitrary scale and level for y_t (standardization). Alternatively, a mean square error norm (MSE) can be substituted for the target correlation, as noted in Wildi (2025). In this case, the classic MSE

[13] Of course, when z_t is GDP growth, sign changes are indicative of expansions and recessions. Formally, the precise relationship between the expected duration between sign changes and the lag-one autocorrelation of the predictor is derived under the assumption of Gaussian time series. However, Wildi (2024) demonstrates that this relationship remains robust even when deviations from Gaussianity occur (central limit theorem applied to lowpass-filtered series).

predictor $y_{t,MSE}$ is obtained as a solution to the SSA criterion by insertion of $1/HT_1 := 1/HT_{MSE}$ in the constraint, where HT_{MSE} represents the holding time of $y_{t,MSE}$.

Traditional predictors often yield an excessive number of false sign changes due to leakage of high-frequency components (high-frequency leakage) in the predicted series, as discussed above (see Figure 7). This noisiness is particularly problematic near the onset or termination of economic crises (recessions), where the target fluctuates around zero; reliable tracking of sign changes would facilitate real-time assessment of the occurrence and timing of such events. Wildi (2025) shows that the SSA criterion helps control this phenomenon (see Fig.12 for illustration).
Using the dual formulation of the criterion (4), the author demonstrates that the SSA solution yields the *lowest* zero-crossing rate among all (linear) predictors with equivalent target correlation. The concept can be extended to a multivariate framework, denoted M-SSA, by generalizing the correlation functions as detailed in the next section. As we shall see, controlling the zero-crossing rate within the M-SSA framework will have a quantifiable effect on the hit rate and false positive rate, thereby explicitly addressing the detection of present (nowcasting) and future (forecasting) economic contractions and expansions.

We now briefly elucidate the SSA optimization criterion outlined in Eq. (2). Consider a simplified univariate framework where $x_t = \epsilon_t$ represents white noise with variance σ^2 and assume, for notational simplicity, that $\sigma^2 = 1$. A generic target specification for z_t is then defined as a linear filter of x_t:

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

where it is assumed that $\gamma_k \neq 0$ for some $k < 0$ (implying an acausal filter), and $\sum_{|k|<\infty} \gamma_k^2 < \infty$ (squared summability), ensuring the stationarity of the target z_t. With these definitions, all processes are centered at zero, justifying the use of Equation (4) which links first-order ACF and HT. Extensions to non-stationary integrated processes are discussed in Wildi (2025). Suppose the goal is to predict z_{t+h}, $h \geq 0$. Let $y_{ht} = \sum_{k=0}^{L-1} b_{hk} x_{t-k}$ denote a causal predictor of length L. For notational convenience we now omit the subscript h from the predictor and we denote $\boldsymbol{\gamma}_h :== (\gamma_h, \gamma_{h+1}, ..., \gamma_{h+L-1})'$. Then, under the postulated white noise assumption we obtain:

$$\rho(z, y) = \frac{\mathbf{b}'\boldsymbol{\gamma}_h}{\sqrt{\mathbf{b}'\mathbf{b}}\sqrt{\boldsymbol{\gamma}_\infty'\boldsymbol{\gamma}_\infty}} \text{ and } \rho(y, 1) = \frac{\mathbf{b}'\mathbf{M}\mathbf{b}}{\mathbf{b}'\mathbf{b}},$$

where $\boldsymbol{\gamma}_\infty'\boldsymbol{\gamma}_\infty := \sum_{|k|<\infty} \gamma_k^2 < \infty$ and where the so-called autocovariance generating matrix $\mathbf{M}$, of dimension $L \times L$, is defined as:

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

The matrix $\mathbf{M}$ is such that $\mathbf{b}'\mathbf{M}\mathbf{b} = \sum_{k=1}^{L-1} b_{k-1}b_k$, representing the first-order autocovariance of y_t. Using this, the optimization criterion (2) can be reformulated as:

$$\left.\begin{array}{c} \max_{\mathbf{b}} \mathbf{b}'\boldsymbol{\gamma}_h \\ \text{s.t.} \quad \mathbf{b}'\mathbf{M}\mathbf{b} = l\rho_1 \\ \mathbf{b}'\mathbf{b} = l \end{array}\right\}, \tag{5}$$

where the additional length constraint $\mathbf{b}'\mathbf{b} = l$ serves two purposes: firstly, to ensure the uniqueness of the SSA solution; secondly, to normalize the objective function $\mathbf{b}'\boldsymbol{\gamma}_h$, making it proportional to the target correlation $\rho(z, y)$, up to a fixed scaling factor $\sqrt{\mathbf{b}'\mathbf{b}}\sqrt{\boldsymbol{\gamma}_\infty'\boldsymbol{\gamma}_\infty} = \sqrt{l}\sqrt{\boldsymbol{\gamma}_\infty'\boldsymbol{\gamma}_\infty}$ (for convenience we often select $l = 1$, leading to a standardized y_t). Criterion (5) is readily amenable to optimization, as it involves affine and quadratic functions of $\mathbf{b}$, see Wildi (2024) for background (detailed derivations are obtained in Wildi (2025)).
Under the above assumptions, the classical MSE predictor of length L can be expressed via orthogonal projection as:

$$y_{MSE,t} = \sum_{k=0}^{L-1} \gamma_{k+h} x_{t-k}$$

which corresponds to setting $b_{MSE,k} = \gamma_{k+h}$, $k = 0, ..., L-1$. In general, to enhance the smoothness of the SSA predictor y_t, we select $\rho_1 > \rho_{MSE}$ in criterion (5), where ρ_{MSE} denotes the first-order ACF of the (benchmark) MSE predictor[14]. Finally, an extension to autocorrelated $x_t \neq \epsilon_t$ is discussed below.

4.2 M-SSA Optimization

We can now extend the SSA criterion (5) to a multivariate setting. Let $\mathbf{X}_t$ (of dimension $t \times n$) denote a set of n explanatory series $\mathbf{x}_{1t}, ..., \mathbf{x}_{nt}$, with observations $\mathbf{x}_{it} = (x_{it}, ..., x_{i1})'$. Let $\mathbf{Z}_t$, of dimension $t \times m$, represent a set of m target series $\mathbf{z}_{it} = (z_{it}, ..., z_{i1})'$, $i = 1, ...m$, which typically lie outside the linear space spanned by $\mathbf{X}_t$ (and are unknown at time t but not independent of $\mathbf{X}_t$). In this framework, $\mathbf{X}_t$ represents a matrix of selected economic indicators for the German economy, with $n = 5$. The targets z_{it}, i=1,...,5, are obtained by applying a two-sided HP(160) filter to the doubly-infinite extensions x_{ik}, $k \in \mathbf{Z}$, of these indicators. In a subsequent step, as delineated in Section (5), the (filtered) predictors y_{hit} of z_{it+h}, with $h \geq 0$, will be utilized to construct predictors for the GDP. However, unlike the approach in Section (3), these predictors will be based on multivariate M-SSA filters rather than univariate HP-C designs.

For this purpose, to introduce the M-SSA framework, we begin by assuming that $\mathbf{X}_t = \boldsymbol{\epsilon}_t$ is an n-dimensional white noise sequence with a full-rank variance-covariance matrix $\boldsymbol{\Sigma}$. Consider the generalized target $\mathbf{z}_t$:

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

[14] Selecting $\rho_1 < \rho_{MSE}$ would produce an SSA predictor with more frequent zero-crossings than the MSE benchmark, which is typically undesirable in practical applications.

where $\mathbf{\Gamma}_k$ is a sequence of filter matrices of dimension $n \times n$ (hence we assume $m = n$), satisfying the square-summability condition. Each $\mathbf{\Gamma}_k$ has entries (γ_{ijk}) for $i, j = 1, \ldots, n$. We focus on the estimation of $\mathbf{z}_{t+h}$, utilizing the $n-$dimensional predictor defined by

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

with $n \times n$-dimensional filter matrices $\mathbf{B}_k$, $k = 0, ..., L-1$ (for notational simplicity we drop the subscript h referring to the forecast horizon in these expressions). Let b_{ijk} denote the entries of the matrix $\mathbf{B}_k$ for $k = 0, \ldots, L-1$. We define several vector notations as follows:

$$\begin{aligned}
\boldsymbol{\epsilon}_{it} = (\epsilon_{it}, \epsilon_{it-1}, ..., \epsilon_{it-(L-1)})' \quad &, \quad \boldsymbol{\epsilon}_{\cdot t} = (\boldsymbol{\epsilon}'_{1t}, ..., \boldsymbol{\epsilon}'_{nt})' \\
\boldsymbol{\gamma}_{ijh} = (\gamma_{ijh}, \gamma_{ijh+1}, ..., \gamma_{ijh+L-1})' \quad &, \quad \boldsymbol{\gamma}_{i\cdot h} = (\boldsymbol{\gamma}'_{i1h}, \boldsymbol{\gamma}'_{i2h}, ..., \boldsymbol{\gamma}'_{inh})' \\
\mathbf{b}_{ij} = (b_{ij0}, b_{ij1}, ..., b_{ijL-1})' \quad &, \quad \mathbf{b}_i = (\mathbf{b}'_{i1}, \mathbf{b}'_{i2}, ..., \mathbf{b}'_{in})' \\
\boldsymbol{\gamma}_h = (\boldsymbol{\gamma}_{1\cdot h}, \boldsymbol{\gamma}_{2\cdot h}, ..., \boldsymbol{\gamma}_{n\cdot h}) \quad &, \quad \mathbf{b} = (\mathbf{b}_1, ..., \mathbf{b}_n).
\end{aligned}$$

Here, $\boldsymbol{\gamma}_h$ (or $\mathbf{b}$) is a matrix of dimension $(L \cdot n) \times n$. The i-th column, $i = 1, ..., n$, contains the filter weights associated with the i-th target (or i-th predictor), and the filter weights corresponding to the $j = 1, ..., n$ series (indicators) are stacked into this column vector, of total length $n \cdot L$. The i-th column vector of $\boldsymbol{\gamma}_h$ (or $\mathbf{b}$) can be applied to the $L \cdot n$-dimensional $\boldsymbol{\epsilon}_{\cdot t}$, which stacks all n (artificial) 'indicators' into a single long data vector. Specifically, we define

$$y_{ijt} := \mathbf{b}'_{ij} \boldsymbol{\epsilon}_{jt},$$

which allows us to express the i-th predictor as:

$$y_{it} = \mathbf{b}'_i \boldsymbol{\epsilon}_{\cdot t} = \sum_{j=1}^{n} y_{ijt}.$$

Similarly, the multivariate predictor vector can be written as:

$$\mathbf{y}_t = \mathbf{b}' \boldsymbol{\epsilon}_{\cdot t}.$$

Also, under the white noise assumption, the orthogonal projection leads to the MSE predictor expressed as:

$$\hat{\mathbf{y}}_{t,MSE} = \boldsymbol{\gamma}'_h \boldsymbol{\epsilon}_{\cdot t}.$$

Note that we here assume the MSE predictor to rely on sub-filters of length L for each subseries $\mathbf{x}_{it}$. Recall from the SSA criterion (5) that it is not necessary to rely on the doubly infinite representations of the target when formulating the optimization criterion, a point that is subsequently confirmed. We define the Kronecker products:

$$\tilde{\mathbf{I}} := \mathbf{\Sigma} \otimes \mathbf{I}_{L \times L}, \quad \text{and } \tilde{\mathbf{M}} := \mathbf{\Sigma} \otimes \mathbf{M},$$

where $\mathbf{\Sigma}$ is the variance-covariance matrix of the (white noise) data, $\mathbf{I}_{L\times L}$ is the $L \times L$ identity matrix and $\mathbf{M}$ is the autocovariance generating matrix introduced earlier. Under the simplifying standardized white noise assumption, we can express the following expectations:

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

Consequently, we propose the following multivariate M-SSA criterion for $i = 1, ..., n$:

$$\begin{aligned} &\max_{\mathbf{b}_i} \boldsymbol{\gamma}'_{i\cdot h}\tilde{\mathbf{I}}\mathbf{b}_i \\ &\text{s.t. } \mathbf{b}'_i\tilde{\mathbf{M}}\mathbf{b}_i = \rho_i \\ &\mathbf{b}'_i\tilde{\mathbf{I}}\mathbf{b}_i = 1, \end{aligned} \tag{6}$$

for $i = 1, ..., n$, where we assume an arbitrary normalization, $l = 1$, in the length constraint $\mathbf{b}'_i\tilde{\mathbf{I}}\mathbf{b}_i = 1$. If explicitly necessary, the arbitrary scaling $l = 1$ can be revised and optimized separately at a later stage, after an optimal 'unity-length' M-SSA predictor has been determined. In analogy to the univariate case, the objective function is proportional to the target correlation $\rho(z_i, y_i)$ between the i-th predictor and the i-th target. Moreover, the criterion involves affine and quadratic forms of $\mathbf{b}_i$, which simplifies the process of numerical optimization. Finally, the HT constraint $\mathbf{b}'_i\tilde{\mathbf{M}}\mathbf{b}_i = \rho_i$, where we used $l = 1$, serves to control the first-order ACF or, equivalently, the HT of the predictor.

Finally, we can relax the WN hypothesis and assume a stationary process

$$\mathbf{X}_t = \sum_{k=0}^{\infty} \mathbf{\Xi}_k \boldsymbol{\epsilon}_{t-k}, \tag{7}$$

where $\mathbf{\Xi}_0 = \mathbf{I}$. Then, the target and predictor can be expressed formally as:

$$\begin{aligned} \mathbf{z}_t &= \sum_{|k|<\infty} (\mathbf{\Gamma} \cdot \mathbf{\Xi})_k \boldsymbol{\epsilon}_{t-k} \\ \mathbf{y}_t &= \sum_{j\geq 0} (\mathbf{B} \cdot \mathbf{\Xi})_j \boldsymbol{\epsilon}_{t-j}. \end{aligned}$$

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

$$\mathbf{B}_k = \begin{cases} (\mathbf{B}\cdot\mathbf{\Xi})_0 \mathbf{\Xi}_0^{-1} & , k = 0 \\ \left((\mathbf{B}\cdot\mathbf{\Xi})_k - \sum_{m=0}^{k-1} \mathbf{B}_m \mathbf{\Xi}_{j-m}\right) \mathbf{\Xi}_0^{-1} & , k = 1, ..., L-1 \end{cases} \tag{8}$$

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

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

We now assume $h \geq 0$[15] and L to be sufficiently large to render $(\mathbf{\Gamma} \cdot \mathbf{\Xi})_{h+k} \approx 0$ and $(\mathbf{B} \cdot \mathbf{\Xi})_j \approx 0$ for $k, j \geq L$. We define:

$$\begin{aligned}
(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{ijh} &= ((\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ijh}, (\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ijh+1} ..., (\mathbf{\Gamma} \cdot \mathbf{\Xi})_{ijh+L-1})' \\
(\boldsymbol{\gamma} \cdot \boldsymbol{\xi})_{ih} &= \left((\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i1h}, (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{i2h}, ..., (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{inh}\right)' \\
(\mathbf{b} \cdot \boldsymbol{\xi})_{ij} &= ((\mathbf{B} \cdot \mathbf{\Xi})_{ij0}, (\mathbf{B} \cdot \mathbf{\Xi})_{ij1}, ..., (\mathbf{B} \cdot \mathbf{\Xi})_{ijL-1})' \\
(\mathbf{b} \cdot \boldsymbol{\xi})_{i} &= \left((\mathbf{b} \cdot \boldsymbol{\xi})'_{i1}, (\mathbf{b} \cdot \boldsymbol{\xi})'_{i2}, ..., (\mathbf{b} \cdot \boldsymbol{\xi})'_{in}\right)'
\end{aligned}$$

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

$$\begin{aligned}
&\max_{(\mathbf{b} \cdot \boldsymbol{\xi})_i} (\boldsymbol{\gamma} \cdot \boldsymbol{\xi})'_{ih} \tilde{\mathbf{I}} (\mathbf{b} \cdot \boldsymbol{\xi})_i \\
\text{s.t. } &(\mathbf{b} \cdot \boldsymbol{\xi})'_i \tilde{\mathbf{M}} (\mathbf{b} \cdot \boldsymbol{\xi})_i = \rho_i \\
&(\mathbf{b} \cdot \boldsymbol{\xi})'_i \tilde{\mathbf{I}} (\mathbf{b} \cdot \boldsymbol{\xi})_i = 1
\end{aligned} \tag{9}$$

for $i = 1, ..., n$. In analogy to the above white noise case, the criterion involves affine and quadratic functions of $(\mathbf{b} \cdot \boldsymbol{\xi})_i$ and thus can be solved for the convolution weights. The requested M-SSA coefficients $\mathbf{b}_i$ can subsequently be determined through deconvolution, as indicated in Equation (8). The present discussion is confined to stationary processes, omitting a lengthier treatment of rank-defficient designs (cointegration) in the non-stationary case.

The subsequent analysis employs the M-SSA framework to construct 'smooth' predictors for forecasting z_{it+h} as well as for forecasting the original (unfiltered) GDP. In particular, it is shown that controlling the zero-crossing rate within the M-SSA framework has a quantifiable effect on the hit rate and false positive rate.

4.3 Example: Out of sample M-MSE vs. M-SSA Nowcasts

For illustration, consider the problem of nowcasting HP-filtered GDP, denoted by HP-GDP, where the target z_t is generated using the two-sided HP(160) filter examined previously. We compare the classic multivariate mean squared error (MSE) predictor (denoted M-MSE) with the M-SSA predictor where we set $HT_1 = 1.5 \times HT_{MSE}$. This constraint ensures that the M-SSA predictor produces roughly 33% fewer sign changes than M-MSE in large samples. Comparisons with M-MSE allow us to see the cost of this constraint in terms of prediction performance (target correlation). Both predictors use the same underlying VAR(1) model introduced in Section (3) and estimated on data up to Q4-2007. This ensures a lengthy out-of-sample period that includes significant events such as the financial crisis, the sovereign debt crisis, and the COVID-19 pandemic.[16]

[15] Similar but slightly more complex expressions would arise for $h < 0$ (backcasting), which are omitted here.

[16] The economic sentiment indicator (ESI) does not receive any weight in the parsimonious VAR(1) model estimated over the shorter sub-sample ending in Q4-2007. Technical details and background information on the M-SSA are provided in Wildi (2025). All computations are conducted using the M-SSA package, as documented in Wildi (2025). The package can be accessed at `https://github.com/wiaidp/R-package-SSA-Predictor`.

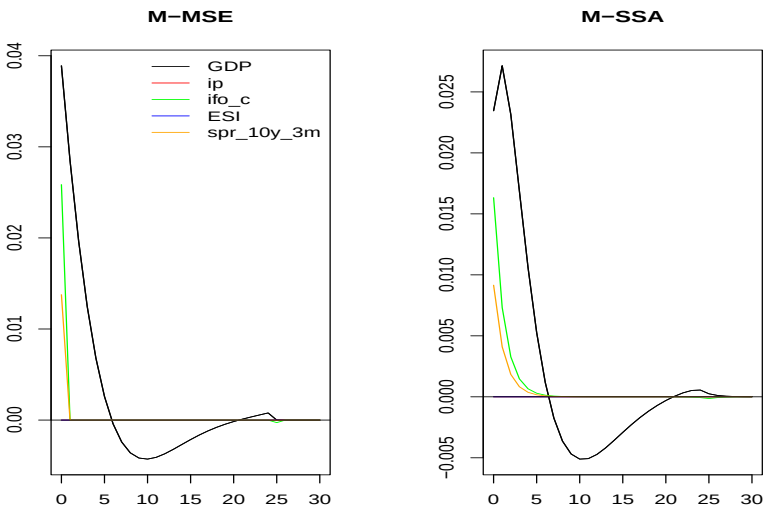


Figure 10: Multivariate M-MSE (left) and M-SSA (right) nowcast filter weights.
Note difference in Scales.

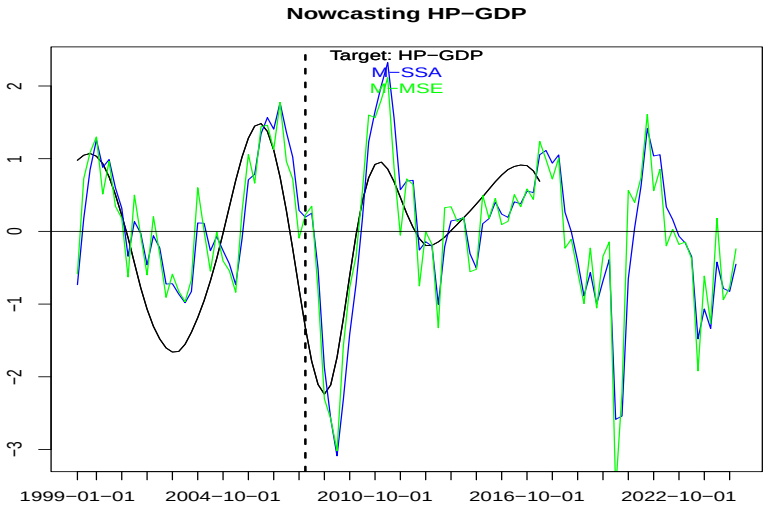


Figure 11: Nowcasts of Smoothed GDP
Black: Target (Two-sided HP(160) applied to GDP)
Blue: M-SSA Nowcast
Green: M-MSE Nowcast
The dashed vertical line delimits in- and out-of sample periods.
Note that the two-sided filter does not extend to the sample end.

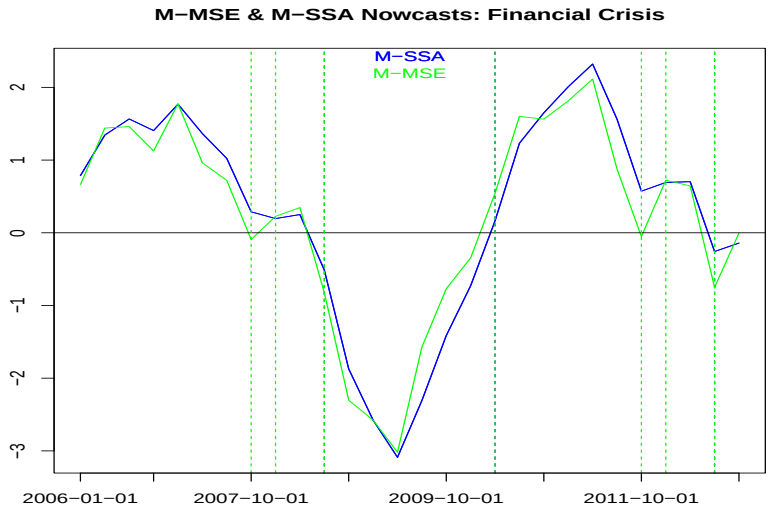


Figure 12: M-MSE (green) and M-SSA (blue) nowcasts around the financial crisis.
Sign changes of the M-MSE marked by green vertical lines.
The holding time (HT) of each filter corresponds to the expected duration between its sign-changes. Here, we impose $HT_{MSSA} = 1.5 \times HT_{MSE}$. In large samples, we expect the M-SSA to exhibit approximately 33% fewer sign changes than the M-MSE.

Table 4: Sample target correlations and holding times (HT) of M-SSA and M-MSE nowcasts.

	M-SSA	M-MSE
Target correlation	0.68	0.69
HT	6.50	3.71

Figure (10) displays the coefficients of the multivariate nowcasts, where M-SSA's increased smoothing reduces the maximum amplitudes of the filter weights and slows their rate of decay at longer lags. The resulting filtered series, presented in Figure (11), are standardized to enable direct comparisons. The M-SSA nowcast is slightly smoother than that of the M-MSE, exhibiting approximately one-third fewer sign-changes. Fig.(12) compares both nowcasts in relation to the financial crisis, highlighting that the classic (MSE-based) design tends to produce unwanted 'noisy' zero-crossings at (or in the vicinity of) the transitions between recessions and expansions. Further, Table (4) compares sample target correlations and HTs of the two nowcasts. Increased smoothness, i.e. sample holding time increases from 3.7 to 6.5 quarters, achieved by the M-SSA comes at the cost of a slight reduction (0.68 vs 0.69) in target correlation. This highlights the inherent trade-off—often referred to as the prediction dilemma—between smoothness (zero-crossing rate) and predictive accuracy (target correlation) within the M-SSA criterion. The results confirm that the smoothing constraint in M-SSA increases by 50% the HT relative to that of the M-MSE predictor[17] but has only a modest effect on the target correlation, suggesting that this trade-off could be beneficial in our application .

[17] The convergence of the sample HT to the effectively imposed HT_1 can be verified in a simulation study, based on very long samples of the time series, utilizing the M-SSA package.

4.4 Example: M-SSA Forecasts

Next, we consider forecasts of HP-GDP, changing the previous $h = 0$ (nowcast) to $h = 1, ..., 6$ quarters. The influence of the forecast horizon on the M-SSA design is depicted in Figure (13), which compares M-SSA filter weights for the nowcast and the one-year-ahead forecast. Noting the difference in vertical scales, we see that as h increases (right panel), the filter weights diminish, reflecting increased forecast uncertainty. We also see increased weight on the leading indicators relative to that on GDP.[18] Additionally, a phase shift associated with the filter applied to GDP is observed (black line, right panel), reflecting the cyclical characteristics of the HP filter. While this phase effect might cause a sign change in the forecast as h increases — something to be avoided in this context — the low-pass filters assigned to the additional 'leading' indicators (green and yellow lines, right panel) continue to effectively track the low-frequency components of HP-GDP. The combination of the phase effect, applied to GDP, and level-tracking through these additional indicators enables more refined and effective tracking of the target compared to univariate forecasts, which rely solely on the phase-effect of the filter to look ahead.

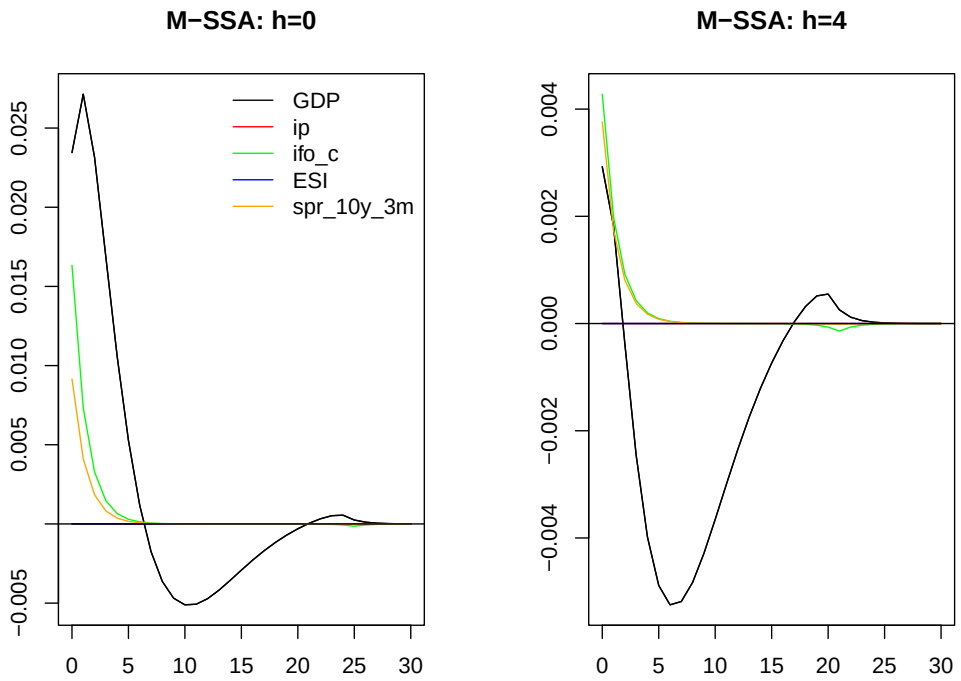


Figure 13: Multivariate M-SSA filter weights
Left: $h = 0$ (nowcast) Right: $h = 4$ (one year forecast)
Note the difference in scales.

M-SSA predictors are compared to the traditional univariate HP-C in Figures (14) and (15). In the first figure, we compare both filters across the periods of the financial crisis (top panels) and the COVID-19 pandemic (bottom panels), for forecast horizons $h = 0, ..., 6$. In the second figure,

[18] Weights on IP and ESI are indistinguishable from zero.

we compare both filters to the forward-shifted GDP for shifts larger than two quarters. All series are standardized for ease of comparison.

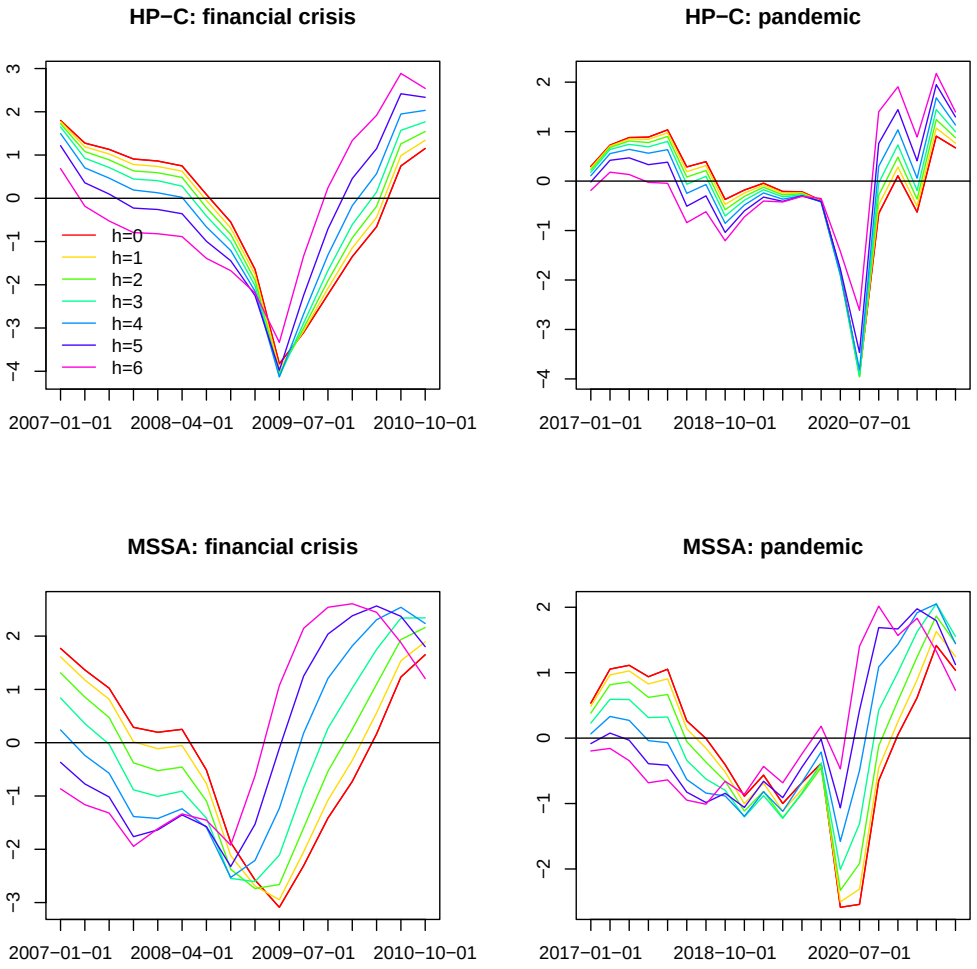


Figure 14: Now- and forecasts of HP-GDP for $h = 0, ..., 6$: HP-C (top) vs. M-SSA (bottom) across the financial crisis (left) and the pandemic (right).

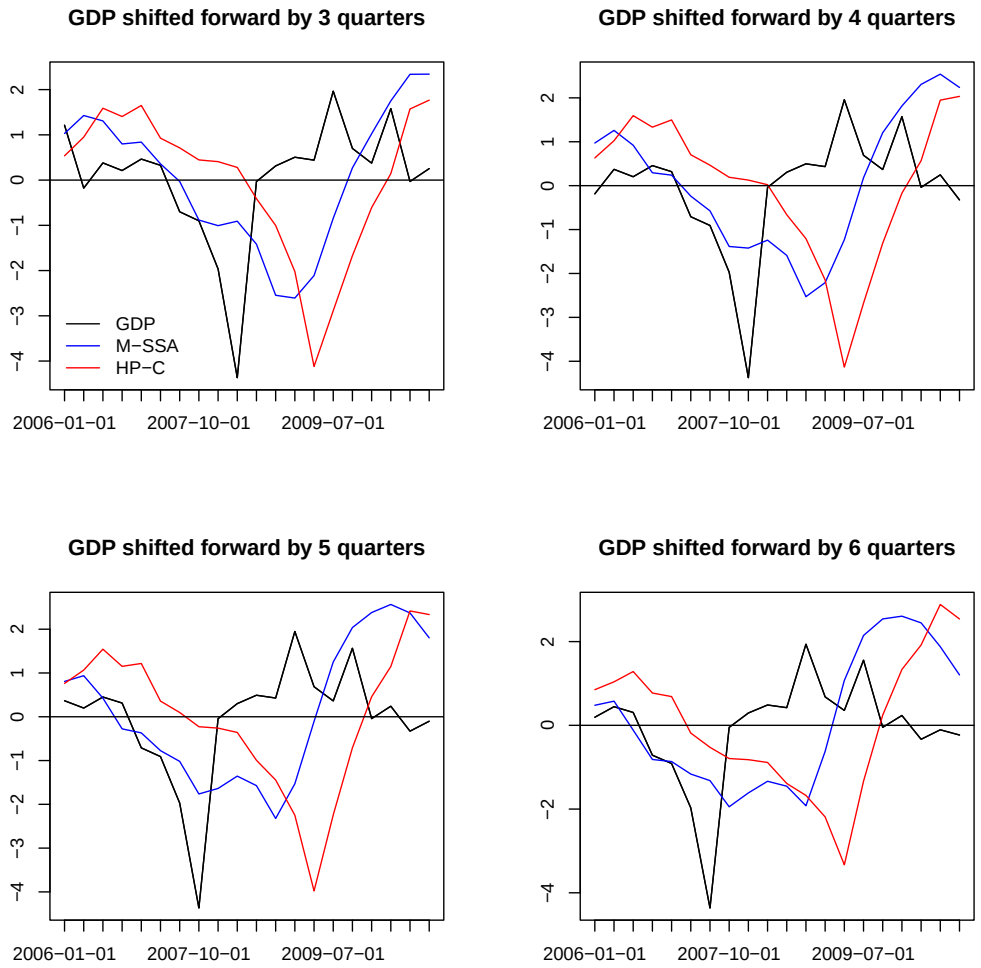


Figure 15: Forward-shifted GDP (black), M-SSA (blue) and HP-C (red) for $h = 3, 4, 5, 6$ across the financial crisis. All series are standardized for ease of comparison.

Forecasts using the M-SSA approach consistently shift leftwards as h increases; this shift appears to be consistent across all levels, including the peaks and troughs of the series. Conversely, the positions of peaks and troughs in the HP-C series are largely unaffected by increases in h. The leftward-shift of the M-SSA forecasts can be partly attributed to the increasing weight assigned to the additional leading indicators, which dominate the GDP component at larger forecast horizons. The systematic left-shift of the M-SSA is a key determinant of its out-of-sample forecast performances, as analyzed in the following section.

5 M-SSA Component Predictor

The targets of the multivariate filters above were the smoothed outputs obtained from the acausal two-sided HP(160) filter applied to the five indicators. However, our objective is to develop predictors for unfiltered GDP. To achieve this, we use the above M-SSA outputs as predictors for forward-shifted GDP.

All the results that we present here exclude the pandemic; results including the pandemic are in the Appendix.

5.1 Tracking GDP: Optimal Weighting

Let $\text{M-SSA}_t^{\text{GDP}}(h)$ denote the M-SSA predictor of HP-GDP for forecast horizon h. A predictor of GDP can be obtained from the simple regression equation

$$\text{GDP}_{t+shift} = \beta_0(shift, h) + \beta_1(shift, h) \cdot \text{M-SSA}_t^{\text{GDP}}(h) + \epsilon_t^{shift,h} \tag{10}$$

Define the (out-of-sample) GDP component predictor $\hat{\text{GDP}}_{t+shift}(h)$:

$$\hat{\text{GDP}}_{t+shift}(h) := \hat{\beta}_0(shift, h, t) + \hat{\beta}_1(shift, h, t) \cdot \text{M-SSA}_t^{\text{GDP}}(h), \tag{11}$$

where the regression parameters $\hat{\beta}_0(shift, h, t), \hat{\beta}_1(shift, h, t)$ are estimated using data up to time point $t-1$ (expanding window). Since M-SSA is based on data up to Q4-2007, $\hat{\text{GDP}}_{t+shift}(h)$ constitutes an out-of-sample predictor of $\text{GDP}_{t+shift}$ for all $shift \geq 0$ and all t larger than Q4-2007. [19] Define the out-of-sample forecast error

$$\hat{\epsilon}_t^{shift,h} := \text{GDP}_{t+shift} - \hat{\text{GDP}}_{t+shift}(h). \tag{12}$$

While we keep the M-SSA outputs fixed, as based on data up to Q4-2007, we implement an expanding window starting from Q1-2008 for deriving the component weights, which are updated on a quarterly basis.

5.2 Relative MSFEs

The statistical significance of the (out of sample) GDP component predictor in Equation (11) can be evaluated by regressing $\text{GDP}_{t+shift}$ on $\hat{\text{GDP}}_{t+shift}(h)$:

$$\text{GDP}_{t+shift} = \alpha_0 + \alpha_1 \hat{\text{GDP}}_{t+shift}(h) + \nu_t^{shift,h}, \tag{13}$$

for observations after Q4-2007: the significance of the corresponding HAC-adjusted test statistics for $H_0 : \alpha_1 = 0$ is indicated by a '*' symbol in Table (5)[20]. The MSE forecast performance of $\hat{\text{GDP}}_{t+shift}(h)$ can be benchmarked against the mean predictor (see Table (5) which reports relative Root Mean-Square Errors, rRMSE) as well as against the direct forecast (see Table (7)). For consistency, both benchmark predictors are also quarterly updated out-of-sample designs.

Our results can be summarized as follows. When excluding the pandemic, the direct forecasts outperform the mean forecast up to two quarters ahead, see Table (8) in the appendix. In contrast, the GDP component predictors $\hat{\text{GDP}}_{t+shift}(h)$, optimized for longer forecast horizons $h \geq 5$,

[19] Equations (10) and (11) could be extended to incorporate additional M-SSA components, but for illustration purposes, we focus on the simplest (and most robust) design.

[20] '*' indicates significance at the 5% level; '**' indicates significance at the 0.1% level. Note that statistical significance is influenced by the relatively short out-of-sample span, of length 17 years.

outperform the mean benchmark up to 5Q ahead, see Table (5). Additionally, $\hat{\text{GDP}}_{t+shift}(h)$, optimized for large forecast horizons $h \geq 4$, surpass the direct forecasts for $shift \geq 3$, see Table (7) in the appendix. Statistical significance of the GDP component predictor, optimized for larger forecast horizons, can be established for at least up to one year ahead, see Table (5). Finally, the strong performance of predictors optimized for larger forecast horizons ($h \geq 4$), at smaller shifts ($shifts < 4$) corroborates the importance of the left-shift of the corresponding predictors, see Figs.(14) and (15), in relation to forecast performance.

Table 5: rRMSEs: out-of-sample forecast errors Equation (12) vs expanding mean of GDP
Out-of-sample evaluation starting in Q1-2008, without pandemic
Values < 1 indicate superior predictions from Equation (11)
Row-index: Shift refers to the left-shift of the target $\text{GDP}_{t+shift}$
Column-index: h refers to the forecast horizon for which $\text{M-SSA}_t^{\text{GDP}}(h)$ is optimized
Significance refers to the hypothesis $\alpha_1 = 0$ in Equation (13).

	h=0	h=1	h=2	h=3	h=4	h=5	h=6
Shift=0	0.975	0.941*	0.886*	0.839**	0.843**	0.897**	0.965
Shift=1	1.085	1.068	1.011	0.927*	0.871**	0.875**	0.919**
Shift=2	1.076	1.089	1.074	1.011	0.938*	0.902**	0.909**
Shift=3	1.037	1.069	1.090	1.059	0.992	0.943*	0.930*
Shift=4	1.013	1.058	1.101	1.097	1.037	0.967	0.925*
Shift=5	1.016	1.062	1.106	1.115	1.078	1.015	0.955*

5.3 Hit Rate vs. False Alarm Rate

To this point, indicators have been assessed by the extent to which they correlate with or lead movements in the target variable. An alternative approach would to be consider their ability to predict the correct "type" of outcome, e.g. whether growth will be stronger or weaker than average, or whether the economy will be in an expansion or a recession. This may be of particular interest to policymakers whose primary concern is to intervene "in the right direction" so as to stabilize the economy overall. We therefore limit our analysis to binary classifiers, focussing in particular on the ROC curve and the AUC.[21]

Consider a binary random variable $Z \in \{0, 1\}$ as our target, and a continuous variable Y. Our predictor for Z is given by the function

$$\mathbf{1}(Y \leq \tau) \equiv I(\tau) \tag{14}$$

where $\mathbf{1}()$ is the indicator function.[22]

The joint distribution of Z and Y allows us to define the Hit Rate $H(\tau)$ and the False Alarm Rate $F(\tau)$ as

[21] For example, see Yang et al. (2024) for a recent discussion of the ROC, the AUC and their relation to selected other binary classifiers. The discussion below follows their notation.

[22] The value of the indicator function is equal to 1 when the condition in its argument is satisfied; otherwise the value of the function is equal to 0.

$$H(\tau) \equiv Pr(Y \leq \tau | Z = 1) \tag{15}$$

$$F(\tau) \equiv Pr(Y \leq \tau | Z = 0) \tag{16}$$

The usefulness of a given variable Y as a predictor will obviously depend on the value of τ selected, which in turn may be influenced by the relative importance attached to $H(\tau)$ and $F(\tau)$. To compare the range of possibilities associated with various choices for Y, we may plot on the unit square $F(\tau)$ vs $H(\tau)$ for all values of τ in the range of Y. This is called the Receiver Operating Characteristic (ROC) curve.

An ideal predictor would be one where $H(\tau) = 1$ and $F(\tau) = 0 \; \forall \tau$ (i.e. $Y > \tau$ whenever $Z = 0$ and $Y \leq \tau$ whenever $Z = 1$.) In this ideal case, the ROC curve would consist of the extreme left and top edges of the plot. Another extreme case would be one where $\{Y, Z\}$ are independent, so $H(\tau) = F(\tau) \; \forall \tau$. In this case, the ROC curve would simply be the 45-degree line through the origin. When comparing two indicators Y_a and Y_b, if the ROC curve of Y_a lies strictly above and to the left of that of Y_b, then we would always prefer to use Y_a as our predictor in place of Y_b since for any choice of τ it will always give us a superior trade-off between $H(\tau)$ and $F(\tau)$. When the ROC curves for two indicators cross, however, then our preferred indicator will depend on the relative importance that we attach to the hit and false alarm rates.[23]

A popular approximate measure of the overall performance of an indicator Y that abstracts from the choice of τ is the Area Under the ROC Curve (AUC). Because the ROC is defined on the unit square, the maximum value of the AUC = 1. Above we saw that when $\{Y, Z\}$ are independent, the ROC curve lies on the 45-degree line, so the area under the curve is then 0.5. The range of AUCs between 1 and 0.5 therefore provides some quantitative guidance as to which predictor variables Y have the potential to improve the prediction of Z.[24].

ROC curves are generated by evaluating the relationship between the sign of the target variable $GDP_{t+shift}$ and events where the forecast $\hat{GDP}_{t+shift}(h)$ exceeds a threshold τ. The threshold τ is varied smoothly across the range of the predictor space to construct the ROC curve.[25] We examined all pairwise combinations of h and $shift$, but for simplicity of exposition, the results below focus on the diagonal case where $h = shift$. For clarity, these results are based on the entire sample, excluding the pandemic period. Fig.(16) compares ROCs of the predictors for small and medium-term forecast horizons and Table (6) presents AUC (area under the curve) for forecast horizons 1-5 quarters ahead.

At forecast horizons less than or equal to two quarters, the direct forecast challenges the filters. However, for horizons $h > 2$, the filters outperform the direct approach: the univariate HP-C design is surpassed by the classic multivariate MSE filter, which is in turn dominated by the M-SSA. These results confirm that controlling the rate of sign changes of the predictor, while maximizing its target

[23] As shown in Wildi (2025), the sign accuracy can be related to the target correlation in the objective function of the (M-)SSA criterion; additionally, the rate of false alarms depends on the holding time (HT), so that the tradeoff underlying the ROC curve is to some extent captured by the M-SSA criterion.

[24] AUCs < 0.5 suggest that $-Y$ may give more useful predictions than Y.

[25] Although we are using standardized GDP, the regressions automatically account for scales and level shifts, so the results directly apply to the original GDP.

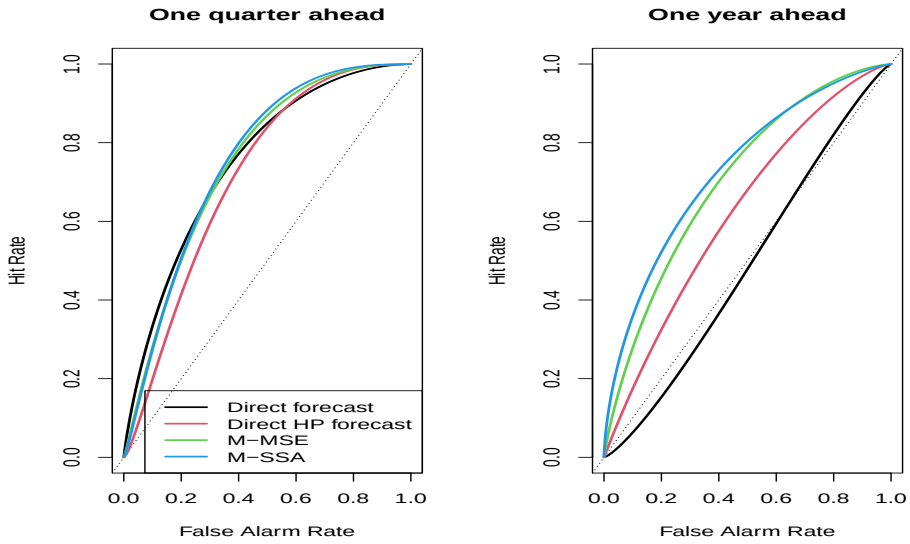


Figure 16: Hit rate vs. false alarm rate of the predictors at small (one quarter ahead:left panel) and yearly forecast horizons (right panel).

Table 6: AUCs of predictors against GDP at various forecast horizons.

	Direct forecast	Direct HP forecast	M-MSE	M-SSA
Forecast horizon 1	0.752	0.715	0.750	0.756
Forecast horizon 2	0.697	0.663	0.714	0.744
Forecast horizon 3	0.554	0.622	0.666	0.716
Forecast horizon 4	0.486	0.624	0.707	0.732
Forecast horizon 5	0.593	0.636	0.660	0.713

correlation through M-SSA can enhance predictor performance in terms of sign accuracy and false alarm rate.

6 Revisions

Revisions refer to changes in the historical values of a time series resulting from updates with new information. Revisions of the M-SSA component predictor can arise from data updates, which we do not analyze further here[26], or from updates to predictor parameters. The latter can be further separated into changes in the regression weights and VAR parameters.

6.1 Revisions: Regression Equation Only

For illustration, we focus on a forecast horizon $h = 4$ for the GDP component predictor and a forward-shift $shift = 4$ for GDP, noting that similar results would be obtained with other combinations of h and $shift$. Figure (17) compares the final and real-time predictors: as the sample size increases, the real-time predictor converges steadily to the final predictor.[27] As a confirmation,

[26] Except for GDP, the other indicators are not or only slightly revised. Moreover, the smoothing effect of filters mitigates the effect of revisions when compared to direct forecasts.

[27] The rate of convergence is partially determined by the simplicity of our design, which avoids incorporating additional M-SSA components.

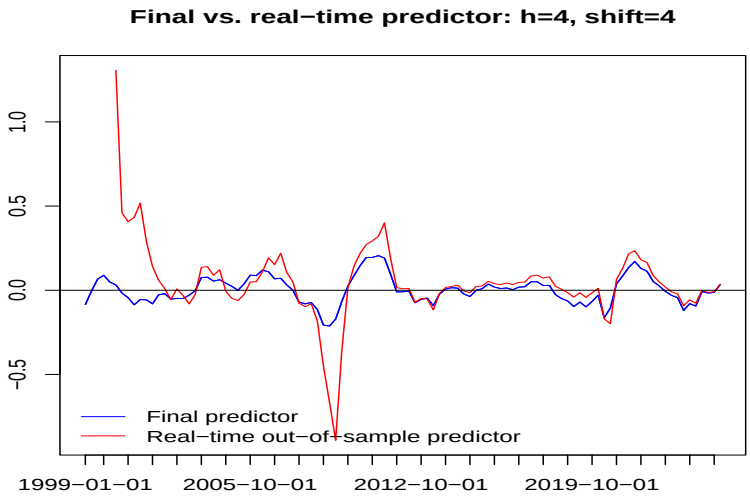


Figure 17: Revisions of the M-SSA GDP component predictor are shown, with the final predictor in blue and the real-time design in red. These revisions result from quarterly updates to the regression coefficients of equation (11). The M-SSA filter itself remains fixed, based on data up to Q4-2007.

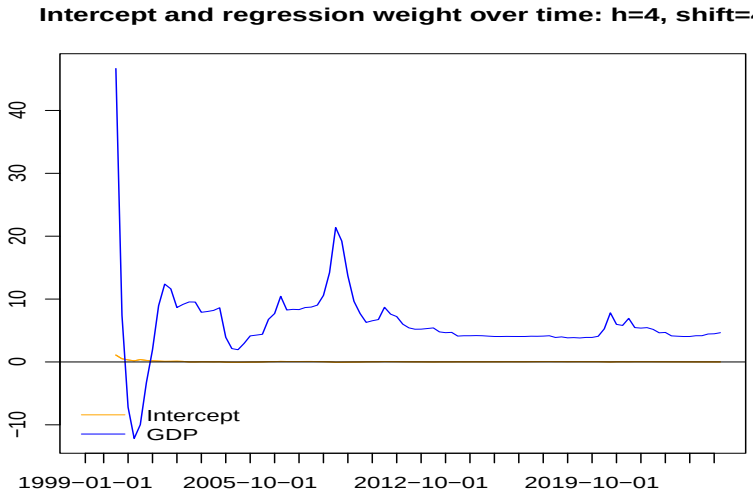


Figure 18: Revisions of intercept and regression weight of Equation (11).

Figure (18) displays the evolution of the intercept and regression weights over time: with increasing sample size, these estimates seem to converge to fixed points, indicating both the stationarity of the process and the consistency of the estimates. In contrast, more complex designs based on multiple M-SSA components can exhibit substantial fluctuations in the regression parameters, sometimes changing signs. This suggests instability and complicates interpretation.

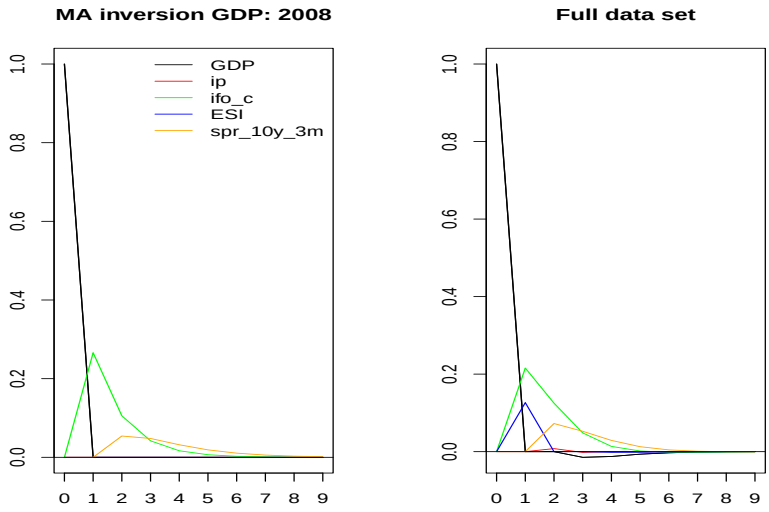


Figure 19: MA-inversion (impulse response) of VAR: data up to Q4-2007 (left) vs. entire data set without pandemic (right).

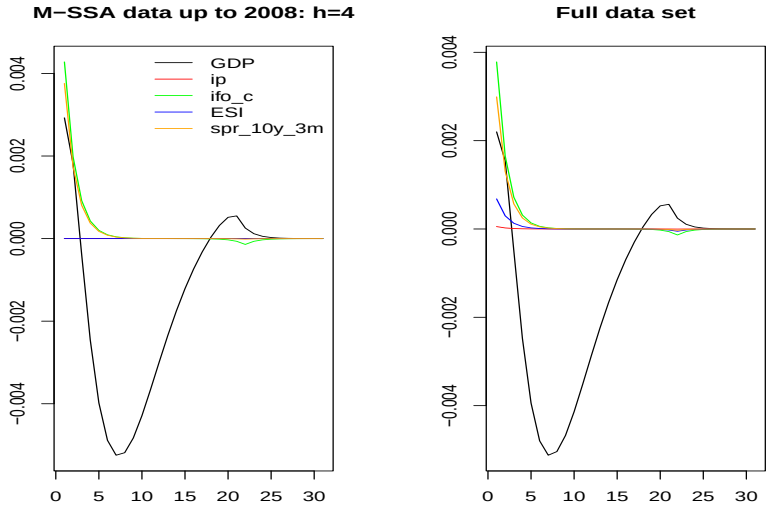


Figure 20: M-SSA filter tracking HP-GDP at forecast horizon $h = 4$: data up to Q4-2007 (left) vs. entire data set without pandemic (right).

6.2 Revisions: M-SSA Only

The second source of revisions in the M-SSA component predictor arises from updates to the VAR(1) model within the M-SSA filter[28]. We compare the original estimates, based on data up to Q4-2007, with the final estimates obtained using $h = 4$ and $shift = 4$ (similar findings apply to other values of forecast horizon and forward-shift). For consistency, we excluded the entire pandemic episode from the data. Figure (19) compares the MA-inversions (impulse responses) of the GDP equation derived from the 'old' and the updated VAR models: unlike the former, the

[28] Similar findings apply to more sophisticated alternatives such as the VAR(3) discussed in Section (2).

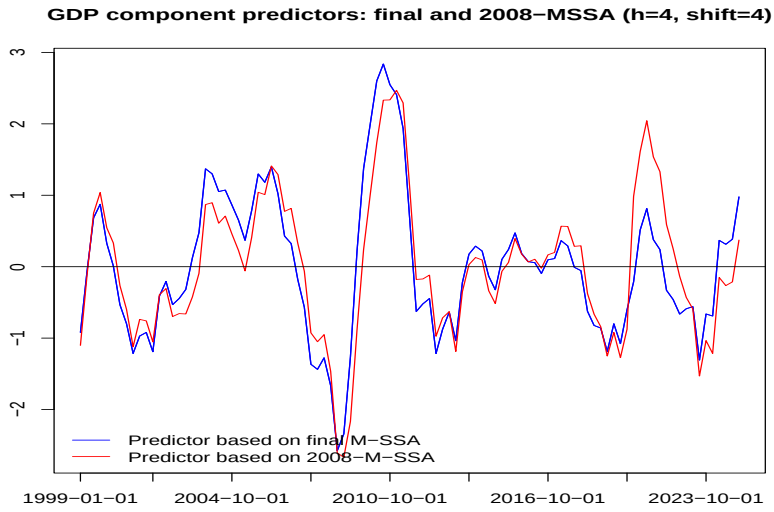


Figure 21: Revisions resulting from updates to the VAR model in M-SSA optimized for $h = shift = 4$. To isolate the effect of the regression revision, all series are standardized. The updated M-SSA (blue) is left-shifted at the zero-crossings and appears slightly faster. For consistency, the entire pandemic episode has been removed.

latter incorporates ESI as an additional explanatory variable for GDP. Figure (20) illustrates the resulting effect on the M-SSA filter targeting HP-GDP at forecast horizon $h = 4$. The updated filter (right panel) assigns more weight to the additional 'leading' indicators relative to GDP. Finally, Figure (21) compares the M-SSA GDP predictors. To isolate the effect of the regression revision, the figure displays standardized series. The update to the VAR model results in a leftward shift of the predictor (represented by the blue line), which appears marginally more rapid compared to the original predictor (depicted by the red line). It is important to note that the smoothness (zero-crossing rate) remains unchanged, though. This suggests that the out-of-sample performance of the M-SSA component predictor may be better than reported in Section (5.2), which is based on the fixed and increasingly outdated version of M-SSA.

7 Summary and Conclusion

Classic direct forecasts of GDP can outperform simple benchmarks, such as the mean, up to two quarters ahead. However, this limited forecast horizon is often insufficient for certain applications. A key challenge with direct forecasts is that the relevant indicators tend to be noisy, which can mask the true signal and lead to overfitting, see Hastie et al. (2009). In particular, the predictor struggles to timely track dips (slowdowns, recessions) or peaks (recoveries, expansions). To address this issue, we apply a HP(160)-filter to smooth the data, with the smoothing parameter $\lambda = 160$ reflecting a more adaptive design than the classic quarterly HP, in line with an intended forecast horizon of up to one year ahead. The resulting predictor tends to outperform the classic direct forecasts in terms of statistical significance, mainly due to a slight left-shift noticeable at sharp recessions such as the Great Financial Crisis or Covid19-Pandemic. However, despite the filtering, the predictor remains somewhat noisy, potentially producing excessive noisy sign changes at (or in

the vicinity of) the transitions between recessions and expansions. This issue is partly due to noise leakage inherent in the one-sided HP filter.

To address these problems, we propose a multivariate extension of the SSA framework based on Wildi (2024). Unlike the univariate HP filter, the M-SSA can leverage cross-sectional information from indicators leading GDP in real time while controlling for the rate of zero-crossings of the predictor. Consequently, the multivariate filter outputs become increasingly left-shifted as the forecast horizon lengthens, allowing dips and peaks of the target series to be tracked more systematically, while reducing the occurrence of noisy zero-crossings. However, since M-SSA does not explicitly target GDP, an additional step is necessary to derive the final GDP predictor.

For illustration, we consider a simple approach based on regressing the single M-SSA GDP output on future GDP, thereby ignoring the other filter outputs. This predictor is intuitively appealing because the future HP-GDP, i.e., the target of M-SSA—is the low-frequency component of the future GDP target. Consequently, a strong link between M-SSA and future HP-GDP also indicates a connection with GDP, even if the statistical significance is obscured by the noise inherent in GDP. Out-of-sample performance suggests that this predictor is statistically significant at horizons exceeding one year. Moreover, designs optimized for larger forecast horizons outperform the mean benchmark at all considered forecast horizons, and the predictor exceeds the performance of direct forecasts—based on unfiltered or HP filtered indicators—at horizons longer than two quarters. An analysis of hit and false alarm rates within the ROC curve confirms that filtering improves predictive performance for forecast horizons exceeding two quarters. Multivariate approaches can further capitalize on leading indicators within the cross-sectional data. Additionally, the AUC is optimized through more precise control of the zero-crossing rate within the M-SSA framework. Finally, an examination of revision errors associated with quarterly updates of the proposed M-SSA-based GDP component predictor indicates that the regression parameters, VAR model, and resulting M-SSA filters stabilize after an initial burn-in period. This stabilization enhances the model's explainability and interpretability.

In conclusion, the new predictor design integrates the traditional direct forecast approach with a novel multivariate filter to target quarterly GDP growth up to one year ahead. This multivariate approach leverages a set of economic indicators leading GDP in real time while controlling the smoothness of the predictor through the zero-crossing rate.

References

Drechsel, K. and Scheufele, R. (2012). The financial crisis from a forecaster's perspective. *Kredit und Kapital*, 45(1):1–26.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Heinisch, K. and Scheufele, R. (2018). Bottom-up or direct? forecasting German GDP in a data-rich environment. *Empirical Economics*, 54(2):705–745.

Heinisch, K. and Scheufele, R. (2019). Should forecasters use real-time data to evaluate leading indicator models for gdp prediction? german evidence. *German Economic Review*, 20(4):170–200.

Lehmann, R. and Reif, M. (2021). Predicting the german economy: Headline survey indices under test. *Journal of Business Cycle Research*, 17(2):215–232.

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.

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

Phillips, P. C. B. and Jin, S. (2021). Business cycles, trend elimination, and the hp filter. *International Economic Review*, 62(2):469–520.

Tsay, R. S. (2013). *Multivariate time series analysis: with R and financial applications*. John Wiley & Sons.

Tsay, R. S., Wood, D., and Lachmann, J. (2022). *MTS: All-Purpose Toolkit for Analyzing Multivariate Time Series (MTS) and Estimating Multivariate Volatility Models*. R package version 1.2.1.

Wildi, M. (2024). Business cycle analysis and zero-crossings of time series: a generalized forecast approach. *Journal of Business Cycle Research*.

Wildi, M. (2025). Sign accuracy, mean-square error and the rate of zero crossings: a generalized forecast approach. mimeo.

Yang, L., Lahiri, K., and Pagan, A. (2024). Getting the roc into sync. *Journal of Business and Economic Statistics*, 42(1):109–121.

Zou, H. and Hastie, T. (2020). *elasticnet: Elastic-Net for Sparse Estimation and Sparse PCA*. R package version 1.3.

8 Appendix

8.1 Additional Performances Results: Without Pandemic

Table 7: Regression of shifted GDP on M-SSA(h) GDP component.
rRMSEs benchmarked against the direct forecast.
Out-of-sample evaluation starting in Q1-2008, without pandemic.

	h=0	h=1	h=2	h=3	h=4	h=5	h=6
Shift=0	1.131	1.092	1.028	0.973	0.977	1.040	1.119
Shift=1	1.295	1.274	1.206	1.106	1.039	1.044	1.097
Shift=2	1.180	1.194	1.178	1.109	1.029	0.990	0.997
Shift=3	1.029	1.061	1.082	1.051	0.984	0.936	0.923
Shift=4	1.012	1.057	1.100	1.096	1.036	0.967	0.924
Shift=5	0.988	1.033	1.076	1.084	1.048	0.987	0.929

Table 8: Direct forecasts.
rRMSEs benchmarked against the expanding mean.
Out-of-sample evaluation starting in Q1-2008, without pandemic.

	Shift=0	Shift=1	Shift=2	Shift=3	Shift=4	Shift=5
rRMSE	0.862	0.838	0.912	1.008	1.001	1.028

8.2 Performances: Including Singular Pandemic Data

Table 9: p-values for $H_0 : \alpha_1 = 0$, HAC-adjusted Wald test, Equation (13).
including the pandemic.

	h=0	h=1	h=2	h=3	h=4	h=5	h=6
Shift=0	0.080	0.042	0.011	0.001	0.000	0.002	0.011
Shift=1	0.403	0.297	0.188	0.095	0.042	0.039	0.103
Shift=2	0.718	0.594	0.359	0.114	0.019	0.009	0.022
Shift=3	0.955	0.950	0.871	0.569	0.158	0.031	0.013
Shift=4	0.593	0.879	0.957	0.893	0.541	0.183	0.080
Shift=5	0.531	0.875	0.971	0.929	0.605	0.239	0.041

Table 10: rRMSEs: out-of-sample forecast errors Equation (12) vs expanding mean of GDP
Out-of-sample evaluation starting in Q1-2008, including the pandemic.
Values < 1 indicate superior predictions from Equation (11).

	h=0	h=1	h=2	h=3	h=4	h=5	h=6
Shift=0	0.999	0.980	0.951	0.925	0.915	0.926	0.955
Shift=1	1.057	1.051	1.028	0.996	0.977	0.978	0.985
Shift=2	1.031	1.035	1.026	0.996	0.964	0.952	0.964
Shift=3	1.022	1.035	1.042	1.026	0.992	0.965	0.953
Shift=4	1.002	1.023	1.044	1.046	1.024	0.999	0.984
Shift=5	1.010	1.032	1.054	1.060	1.043	1.013	0.983

Table 11: Regression of shifted GDP on M-SSA(h) GDP component.
rRMSEs benchmarked against the direct forecast.
Out-of-sample evaluation starting in Q1-2008, including the pandemic.

	h=0	h=1	h=2	h=3	h=4	h=5	h=6
Shift=0	1.230	1.207	1.172	1.139	1.127	1.140	1.177
Shift=1	0.984	0.977	0.956	0.927	0.909	0.910	0.916
Shift=2	1.039	1.044	1.035	1.004	0.972	0.960	0.972
Shift=3	1.046	1.059	1.067	1.050	1.015	0.987	0.975
Shift=4	0.982	1.002	1.023	1.024	1.003	0.978	0.964
Shift=5	0.979	1.001	1.022	1.028	1.012	0.983	0.953

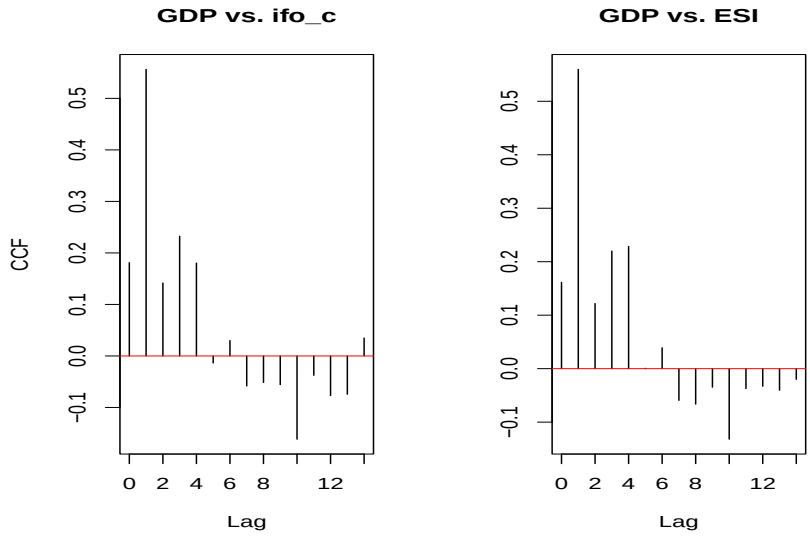


Figure 22: Sample cross-correlation function (CCF): full data set.

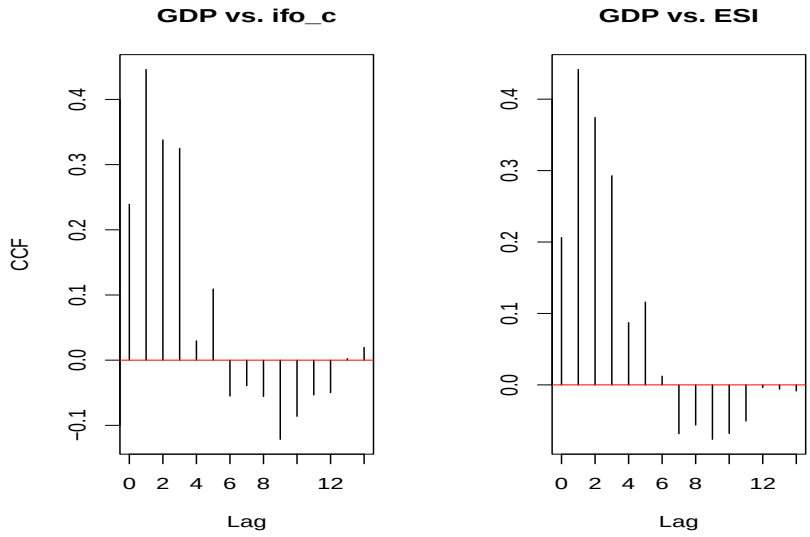


Figure 23: Sample cross-correlation function (CCF): without Pandemic.

Table 12: Direct forecasts.
rRMSEs benchmarked against the expanding mean.
Out-of-sample evaluation starting in Q1-2008, including the pandemic.

	Shift=0	Shift=1	Shift=2	Shift=3	Shift=4	Shift=5
rRMSE	0.812	1.075	0.992	0.977	1.021	1.031