STAFF MEMO

A SMARTer way to forecast

NR. 7 | 2023

FRIDA BOWE
INGA NIELSEN FRIIS
ATLE LONELAND
ERLEND SALVESEN
NJØLSTAD
SARA SKJEGGESTAD
MEYER
KENNETH
SÆTERHAGEN
PAULSEN
ØRJAN ROBSTAD

NORGES BANK

ISSN 1504-2596 (online)
ISBN 978-82-8379-270-6 (online)

A SMARTer way to forecast[*]

Frida Bowe [†] Inga Nielsen Friis[‡] Atle Loneland[§] Erlend Salvesen Njølstad[¶]
Sara Skjeggestad Meyer[‖] Kenneth Sæterhagen Paulsen[**] Ørjan Robstad[††]

28 March 2023

Abstract

In this paper we describe the newly developed System for Model Analysis in Real Time (SMART) used for forecasting and model analysis in Norges Bank. While the long-term goal is to include all empirical models used in forecasting in Norges Bank, the emphasis in this paper will be on the empirical model systems for inflation and GDP. SMART builds on Norges Bank's previous System for Averaging short-term Models (SAM), but with greater flexibility and a richer set of models. In addition, SMART contains a real-time database with a wide-ranging set of historical data, forecasts from empirical models, Norges Bank's forecasts from Monetary Policy Reports (MPR) and forecasts from other institutions (e.g. Statistics Norway). Overall, SMART seems to provide good forecasts and will be a useful tool in the monetary policy process.

[*]The views expressed in this paper are those of the authors and do not necessarily reflect those of Norges Bank. The authors are responsible for any errors and omissions. We would like to acknowledge the essential contributions from Mari Aasgaard Walle, Christer Jahren, Sara Liseth, Michel Hageman, Pål Økern, Claus Erik Kirkerud, Tim Kirstein, Anne Sofie Jore, Mats Elias Davidsen, Maximillian Schröder and Per Bjarne Bye to the development of SMART, and good help from the student interns in Norges Bank. We would like to thank Ole Christian Bech-Moen, Per Espen Lilleås, Karsten Gerdrup and Maximillian Schröder for useful comments on this paper.

[†]Norges Bank. Email: frida.bowe@norges-bank.no.
[‡]Norges Bank. Email: inga-nielsen.friis@norges-bank.no.
[§]Norges Bank. Email: atle.loneland@norges-bank.no.
[¶]Norges Bank. Email: erlend-salvesen.njolstad@norges-bank.no.
[‖]Norges Bank. Email: sara-skjeggestad.meyer@norges-bank.no.
[**]Norges Bank. Email: kenneth.paulsen@norges-bank.no.
[††]Norges Bank. Email: orjan.robstad@norges-bank.no.

1 Introduction

A good understanding of the current state of the economy and how it will evolve over the near future (short-term forecast) is crucial for sound monetary policy decisions. To assess the overall macroeconomic outlook, Norges Bank follows a large set of variables and economic indicators. Some of the variables, such as output, inflation and employment are important target variables in the implementation of monetary policy. Others provide useful information about the economic outlook (e.g. financial market variables, commodity prices, Regional Network and foreign sector variables.).

SMART (System for Model Analysis in Real-Time) is a modern forecasting platform that contains a library of a large set of empirical models and forecasts based on numerous variables. In SMART, model forecasts are averaged based on their real-time forecasting performance, taking into consideration the timing of data releases and revisions in data over time. This follows recent scholarship on economic forecasting, which often finds that a combination of different model provides better forecasts than individual models (Elliott et al. (2006), Koop and Korobilis (2012), Rossi (2013)).

Norges Bank produces most Monetary Policy Report (MPR) forecasts using its core macroeconomic model, the New-Keynesian DSGE model, NEMO (Kravik and Mimir, 2019) (see Figure 1). A core model approach to forecasting is useful for monetary policy analysis and forecast consistency. The forecasts from NEMO reflect the model's theoretical foundation, and Smets and Wouters (2007) show that DSGE models are capable of forecasting reasonably well in the medium run. However, NEMO is not flexible enough to include all relevant information needed to make good short-term forecasts. NEMO includes a limited number of variables, and all variables are at quarterly frequency. Several key model variables are published with a considerable lag and with significant measurement errors. When making short-term forecasts, we want to include a wider set of information about economic developments and take into account more updated information. Forecasts from NEMO are, therefore, conditioned in the short-run on forecasts from a large set of empirical and statistical models, in combination with sector expert judgement.

Furthermore, empirical models are used extensively to cross-check medium and longer-run forecasts from NEMO. The forecasts in the MPR is often a result of iterations between NEMO and cross-check models. This process is important, since we want our forecast to take into account empirical and economic relationships that may not be captured in NEMO.[1] SMART is also used for cross-checks of NEMO-forecasts.

SMART builds on the best models from our previous forecasting system, SAM (Aastveit et al., 2011), but adds a large set of new models. SMART allows for a broader set of models, both in terms of the number of models and model types. In addition, SMART also includes forecasts for inflation and GDP along the full policy horizon, at different frequencies and for sub-components of our variables of interest. To ensure consistency across forecasts, it is possible to restrict monthly and disaggregated forecasts to add up to the quarterly SMART forecasts.

SMART contributes to a more efficient and agile forecasting process at Norges Bank, and frees up time for analytical work. SMART is automated, and by gathering all data and models within a single system, updated model-forecasts are more accessible. The infrastructure allows for fully automated daily updates of the SMART forecasts and updates after important data releases. The flexibility of SMART will speed up and simplify the process of evaluating the quality of new models and comparing them to the existing model system. Historical MPR forecasts are stored in the SMART

[1] One example of this is described in the box "How do interest rates influence household disposable income and consumption?" in Norges Bank (2022).

database, making it easy to evaluate the quality of Norges Bank's monetary policy forecasts over time.

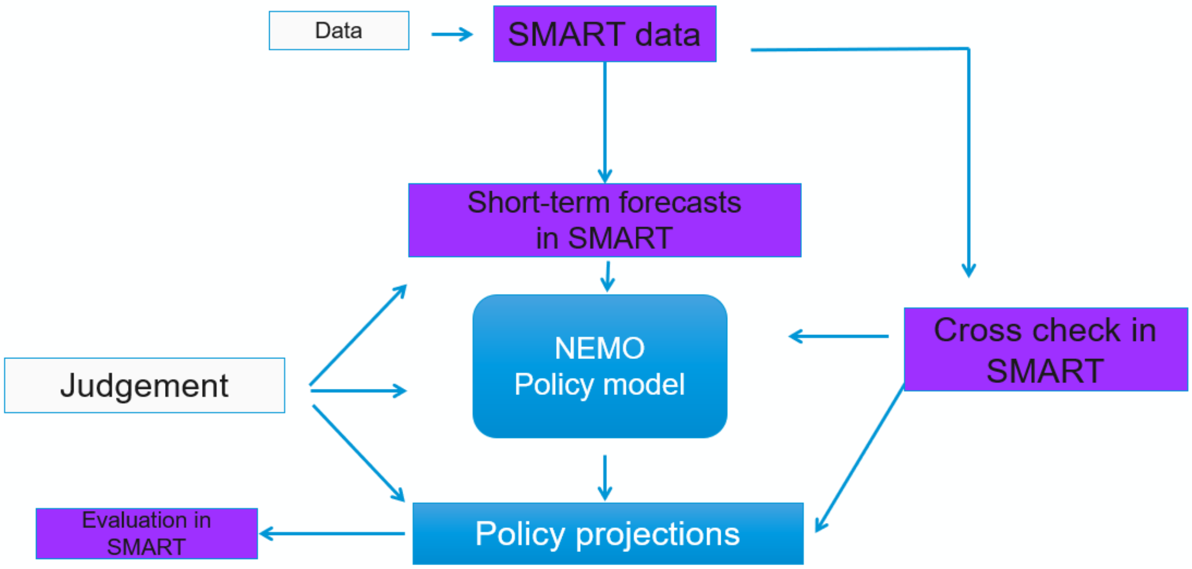


Figure 1: Forecast and policy analysis system (FPAS) at Norges Bank

The system stores all information about the models we use and will act as a real-time model archive for Norges Bank. Estimation results from historical SMART forecasts are available for further analysis. Naturally, some models will change over time, and SMART makes it possible to archive all historical versions of a model.

This paper gives an overview of SMART and is organised as follows: Section 2 presents the infrastructure around SMART, while Section 3 gives an overview of the different data in SMART. In Section 4, we describe the principles applied for combining models into the combined SMART forecast. Sections 5 and 6 document SMART Inflation and SMART GDP. In Section 7, we evaluate the forecasting performance, while Section 8 sums up and presents examples of how SMART forecasts are used in Norges Bank's MPR processes.

2 Overview of the SMART platform

SMART is a platform consisting of several components. Among the most important is the database for storing real-time data and model forecasts, and a Java library for handling the transfer of data to and from the database. The Java library also includes support for some important classes of models[2], as well as good integration with MATLAB and Norges Bank's existing modeling toolbox, NB Toolbox (Paulsen and Loneland, 2021). By supporting a large variety of different model types on one platform that is efficiently and easily accessible by a broad set of users, we lower the cost of including a richer information set in the forecasting process.

[2]A model is here defined as a systematic way of producing forecasts. Examples are empirical models, subjective models, aggregation or combinations. A forecast is a prediction of a future observation of a time series.

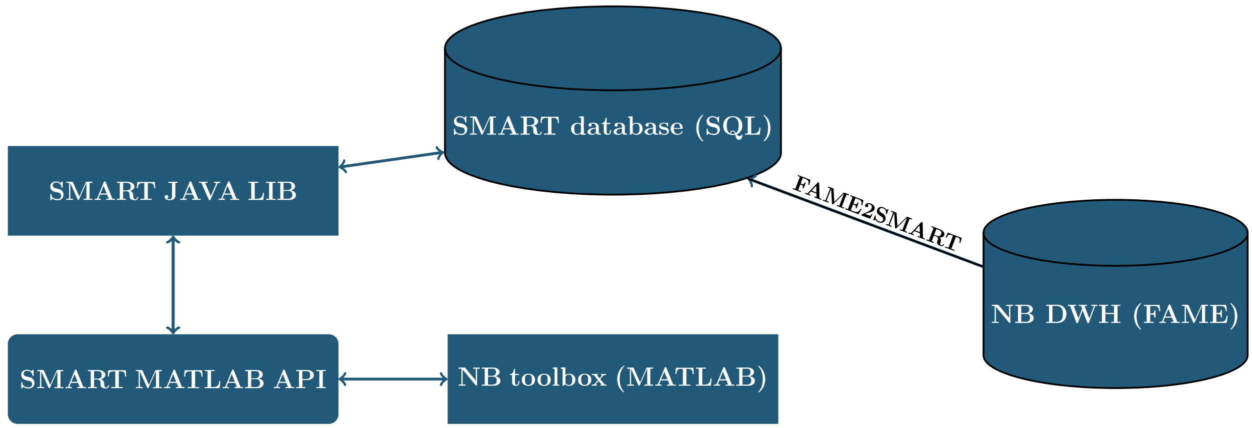


Figure 2: Schematic representation of the SMART system.

Over time, we plan to develop APIs for programming languages other than MATLAB. We are currently working with a version for Python, but the API could easily be extended to the user's preferred analysis package, e.g. Julia or eViews. In addition, more advanced machine learning algorithms will be added, which will leverage big data to develop more sophisticated and, hopefully, useful models.

In the following sections, we provide a introduction to each of the components that make up the SMART system and introduce the most important algorithms employed by the system. See Figure 2 for a schematic outline of the different parts of SMART and how they interrelate.

- **The SMART Database:** An SQL database storing all the metadata, historical data, forecasts, evaluations and scoring of models.
- **SMART JAVA LIB:** A library written in Java for fetching data from and writing forecasts to the SMART database, and for evaluating, combining and aggregating forecasts, in addition to the possibility of converting forecasts from one frequency to another.
- **SMART MATLAB API:** An API written in MATLAB for integrating SMART JAVA LIB into corresponding MATLAB functions and classes. The API mirrors the classes and methods in the SMART JAVA LIB and is fully compatible with Norges Banks's MATLAB toolbox for data management, graphics and econometrics.[3]
- **NB DWH:** Norges Bank's data warehouse based on MarketMap Analytic Platform (previously called FAME).
- **FAME2SMART:** A connector written in .NET/C# for transferring data from Norges Bank's data warehouse.

In addition, SMART comprises a web user interface for easily accessible database search and visualisation and analysis of forecasts, as well as an Excel Add-in for convenient access to historical real-time data and SMART forecast updates.

[3]See Paulsen and Loneland (2021) for more information about this toolbox.

3 The SMART database

3.1 Real-time data

Following the economic forecast literature (Stark and Croushore, 2002; Croushore and Stark, 2003; Clark and McCracken, 2013), all forecasts in SMART are evaluated based on their out-of-sample performance in real time. In the SMART database, real-time data and model forecasts are stored in a timely manner, with vintage tags and with real-time data for each vintage wherever that is available (see Figure 3 for an example). The users of the system are thus able to estimate models and generate forecasts without being at risk of using information that was not readily available to them at the specific time of interest. Currently, 1 638 different variables have been added to the database.

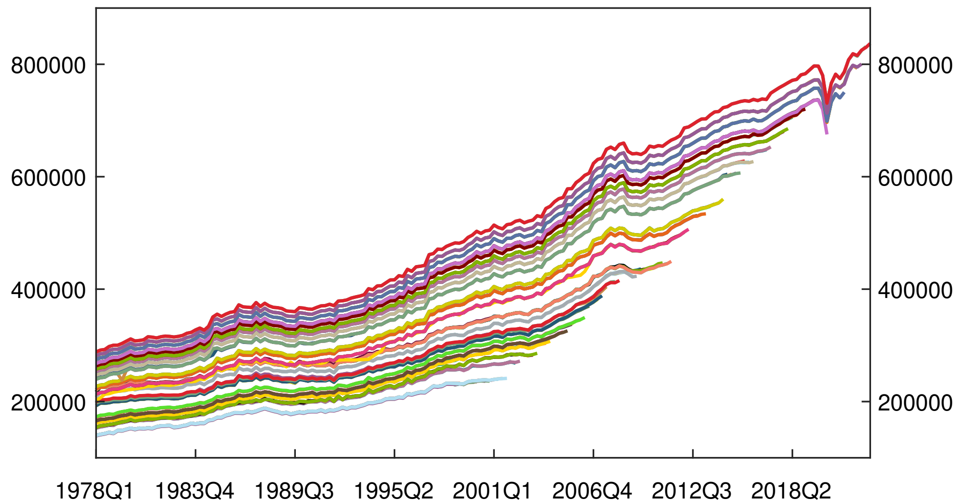


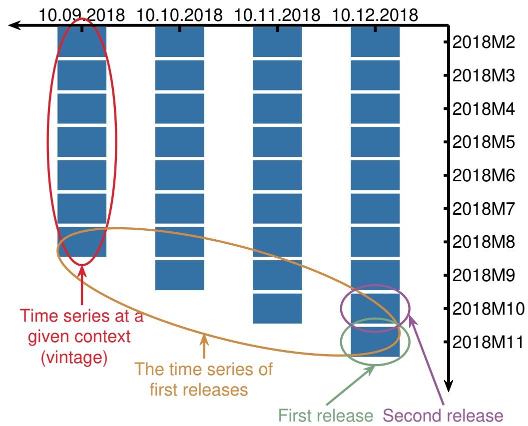
Figure 3: Real-time vintages of real GDP for mainland Norway. Seasonally adjusted. In billions of NOK. 1978Q1-2022Q4

We have defined a vintage in the classical sense, as the full time series at a given publication date. Often the data are revised, and we name the revisions of the observations between vintages for releases. The release one chooses to evaluate the forecasts against is determined by the user and varies from the first release for SMART Inflation to the 12th for SMART GDP. In Figure 4a we illustrate the different data concepts. The vertical blocks highlighted in red refers to the time series as it was at a given publication date. The green and purple circles highlight the first and second release, while the orange covering the diagonal encompasses the time series of first releases. The diagonals above are the time series of second and third releases and so on. If more than one variable is considered, we introduce the concept of context, defined as the latest available data for a set of variables at a given date (Figure 4b). The time stamp, or tag of a vintage or a context, is called a vintage date or a context date, respectively, and the full set of publication dates of all the variables under consideration is named context dates. See the labels on top of Figure 4a and Figure 4b.

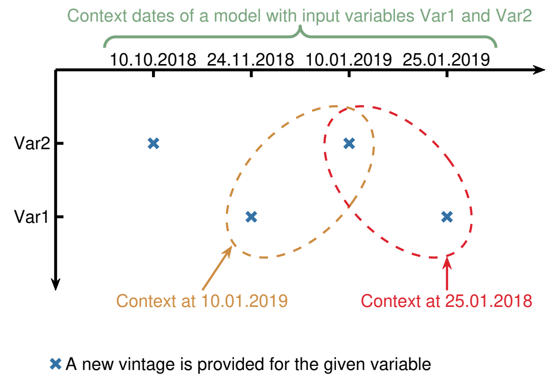
There exist two types of variables that have vintages:

1. The time series is revised back in time and published with a lag. An example of this type is GDP.

2. The time series is not revised back in time, but published with a lag. An example of this type is the CPI.



(a) Real-time illustration of the terms vintages and releases.



(b) Real-time illustration of the contexts of a combined set of vintages of two variables.

If a variable lacks a vintage structure or no vintages exist before a given date, we may construct quasi-vintages. Quasi-vintages are constructed by taking the first vintage available for a variable and removing observations in line with the number of available observations at a given publication date. We estimate the publication dates for these vintages given the publication lag of the first vintage available. A time series that is neither revised, nor published with a lag, is called a series in SMART. An example of a series is the policy rate.

3.2 Calculated objects

In addition to official statistics, many of our models rely on variables that have been produced by Norges Bank staff. These variables are a result of some type of calculation, and we call the results from these calculations for calculated objects. For a formal description of calculations in SMART, see Appendix D.

The purpose of calculated objects is twofold. First, they ensure a separation between official statistics obtained directly from the statistics producer, such as Statistics Norway, and time series that have been remodelled or adjusted by Norges Bank staff. Second, they ensure that all modifications of official data series are recorded and documented for future cross-checks and transparency. All calculated objects are stored in the SMART database in the same manner as a vintage variable or a series.

Common calculated objects consist of aggregations and simple transformations of data series that we use in forecasting. Examples would be (a) calculating the labour share for different consumer sectors, (b) creating a variable consisting of an aggregate, such as income from the household and non-profit sectors, (c) seasonally adjusting data series or (d) adjusting for structural breaks in time series.

Currently, several calculated variables consist of time series that are chained to obtain a longer data history. For instance, Norges Bank's Regional Network was established in 2002, but for many models it is useful to chain the variables with other business and consumer confidence surveys to obtain a longer time series. Key Regional Network variables chained with a synthetic history between 1990 Q1 and 2002 Q2 are included and documented in the SMART database as calculated objects.

Similarly, monthly national accounts (MNA) from Statistics Norway have been available since 2018 with observations starting in January 2016. The short sample creates challenges when setting up robust models. We have therefore estimated a synthetic monthly history for GDP from January 1990 based on a mixed frequency VAR with monthly indicators (see Appendix B).

Moreover, calculated objects are used to estimate new indicators, as for example Norges Bank's high-frequency financial conditions index for Norway (Bowe et al., 2023). The indicators are easily estimated in real-time and used and evaluated alongside other data series in SMART.

To date, over 300 of the 1638 data series in the SMART database are calculated objects. Going forward, we will continue to add relevant real-time data and series to the SMART database. Our ambition is to broaden the variety of data to also include indicators calculated from microdata and unstructured data. During the Covid pandemic, transaction data for Norwegian households rapidly became crucial for understanding and forecasting household behaviour and consumption (Aastveit et al., 2020). Textual data from media articles have also been shown to represent useful high-frequency activity indicators, as in Thorsrud (2018).

3.3 Real-time forecasts

In addition to enabling real-time evaluation of all forecasting models, the SMART database serves as an archive. All indicators and models included in SMART will be stored in the database, together with historical MPR-forecasts and forecasts from other institutions, e.g. Statistics Norway. Henceforth, the SMART database, will archive the data series, forecasting models and model specifications considered in Norges Bank's monetary policy processes at a given time, enhancing transparency. This will also ease future evaluation of Norges Bank's monetary policy forecasts, which are regularly published by the Bank (Norges Bank, 2020).

4 Principles for combining models to SMART forecasts

The primary goal for forecasting in the SMART modelling system is to aggregate information from individual models into a single combined forecast for each variable of interest. The idea of combining or pooling different model-forecasts for the same variable in order to improve the forecasting performance goes back to Bates and Granger (1969). The combined SMART forecast is based on core principles for grouping models into model ensembles, as well as trimming, discounting, and weighting based on historical forecasting performance, see Appendix E.

4.1 Ensembles and hierarchy for models

The SMART system includes a vast number of models, using a mix of different model classes and estimation techniques. Some models are best suited for short term forecasts, while others are expected to perform better at longer horizons. To preserve the differences in model qualities, models in SMART are divided into groups or ensembles (Aastveit et al., 2014), which introduces a multi-step combination hierarchy, as illustrated in Figure 5. At the first level, individual models produce forecasts. At the second level, the forecasts from the individual models are combined into a model ensemble forecast. At the third level, the forecasts from model ensembles at level two are combined into one forecast.

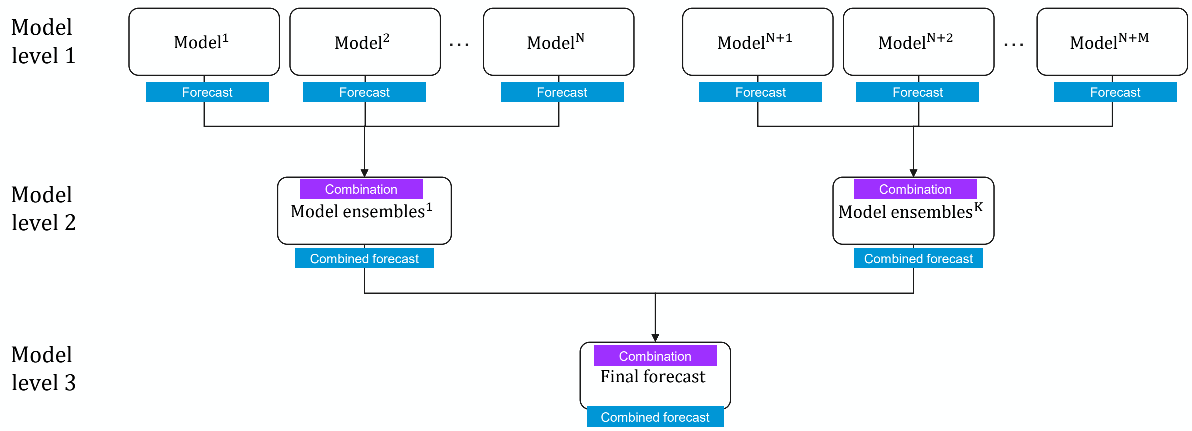


Figure 5: Generic model hierarchy with multiple combination steps

4.2 Measure for forecasting performance

The primary measures for historical forecasting performance in SMART is the out-of-sample Root Mean Square Error (RMSE) for each forecasting horizon.[4] Models are estimated in real-time to account for revisions in the data and data availability, which is important for forecasting macroeconomic variables (Stark and Croushore, 2002).

4.3 Discounting

The historical forecasting performance of models are discounted by use of a discount factor. Discounting tilts the framework in favour of valuing recent forecasting performance over past forecasting performance. This could be important if there is structural changes in the economy and the set of models that provides the best forecasts change over time. One drawback with discounting is that you effectively reduce the evaluation sample, making the model selection more uncertain. The discount factor adjusts the standard calculation of the forecasting performance, $score_m$, by:

$$score_m = \sqrt{\frac{\sum_{i=1}^{N}(x_{m,i} - \hat{x}_i)^2 \times \lambda^i}{\sum_{i=1}^{N}\lambda^i}} \tag{1}$$

where $x_{m,i}$ is the point forecast for period i and model m, $\hat{x}_i$ is the realised value and λ is the discount factor. If λ is equal to 1 there is no discounting. For SMART GDP and SMART Inflation, λ is set at 0.97, resulting in a half-life for forecasting performance after 23 periods (almost 6 years for quarterly models). The discount factor is chosen to ensure a decent trade-off between long-time-performing models and models performing well in the near future. The value of the discount factor may be adjusted in the future.

[4] Other measures such as MSE or MAE are also available in SMART.

To account for cases where models have different starting dates for point forecasts, or where models produce invalid forecasts, models without a valid forecast are given a score relative to the valid models.

Models without a valid forecast are given a score equal to $\widehat{score} + \sigma^P \times W$, where $\widehat{score}$ and σ^P is the mean and standard deviation of the DRMSE for models in the same model ensemble with valid forecasts for the given forecasting date and horizon, respectively. W is a scaling parameter. For SMART GDP and SMART Inflation W is set 1.

4.4 Trimming

At each combination stage in SMART, the number of models included in the weighting process is trimmed. Trimming entails that a subset of the models is given zero weights in the combination, and thus has no impact on the combined forecast. Trimming the number of models included in the credible set of models limits the impact of poorly performing models, and trimming away either poorly performing models or extreme values can improve forecasting accuracy (Jose and Winkler, 2008). The trimming scheme in SMART is flexible and allows for a range of trimming schemes.

SMART GDP and SMART Inflation uses a threshold-based progressive trimming scheme to determine the credible set of models that will be assigned weights in the combined forecast (See Figure 6). This trimming scheme excludes an increasing number of models from the credible set as the number of available models M in a combination stage increases. The threshold levels for the credible set of models Z are given by:

$$Z(M) = \begin{cases} \max(3, 0.5 \times M), & \text{for } 0 < M \leq 10 \\ \max(5, 0.2 \times M), & \text{for } 10 < M \leq 100 \\ \max(20, 0 \times M), & \text{for } 100 < M \end{cases}$$

The number of models included is capped at 20 based on the models' discounted historical performance measure. The values are chosen primarily based on prior experience with model averaging and trimming.

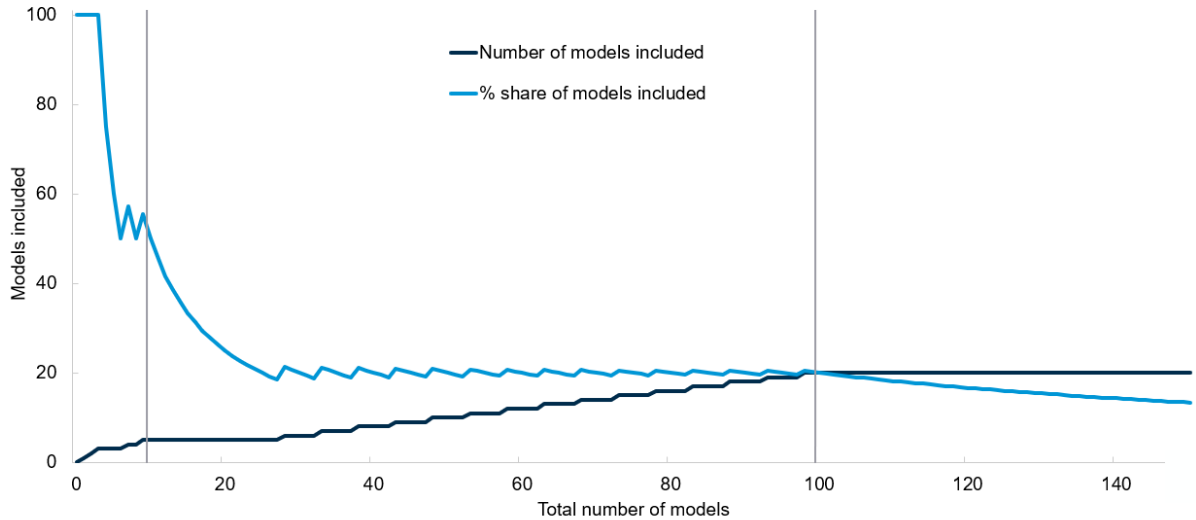


Figure 6: Trimming scheme as number of models increases

4.5 Time-varying weighting

The combined forecast for each combination step will be a weighted average of the subset of models remaining after trimming. These models will be weighted based on the historical performance. The weight assigned to model m can be expressed as:

$$w_m = 1_{\{m \in Z(M)\}} \times \frac{\frac{1}{score_m}}{\sum\limits_{\mu \in Z(M)} \frac{1}{score_\mu}} \tag{2}$$

where Z is the credible set of models based on the trimming scheme, and w_m is positive.

4.6 The combination scheme's impact on performance

While the core principles for discounting, trimming and weighting are based on the forecasting literature (see e.g. Diebold and Shin (2019)), the specific choices are based on judgement and trial and error and may change in future versions of SMART.

We check for robustness along the dimensions summarized in the table below to see how different elements of the combination scheme currently in use affect the forecasting performance of SMART. In addition, we explore the impact on forecasting performance through the quarter from the combining SMART forecasts based on publication dates for MPR. We computed the robustness checks for both model systems, SMART Inflation and SMART GDP.

Model system	Trimming	Time-varying weights	Discounting
Baseline			
SMART	Yes	Yes (RMSE)	Yes (0.97)
Robustness check			
A: No trimming, equal weights	No	No	No
B: No trimming, varying weights	No	Yes (RMSE)	Yes (0.97)
C: Trimming, equal weights	Yes	No	Yes (0.97)
D: Varying discount factor	Yes	Yes (RMSE)	intervals between 0 and 1

Figure 7 summarises robustness check A, B and C, relative to SMART for CPI-ATE (left-hand-side) and GDP (right-hand-side). For CPI-ATE, trimming substantially improves the forecasting performance of the model system at shorter horizons (A). On the other hand, the effect of time-varying weights changes vastly conditional on the trimming scheme: Without trimming, time-varying weights improves the forecasting accuracy (B), but in the case with trimming, weights based on forecasting performance does not have major impact on the forecasting performance (C). The results from robustness checks for GDP are similar to CPI-ATE, but improves forecasting performance to a lesser degree.

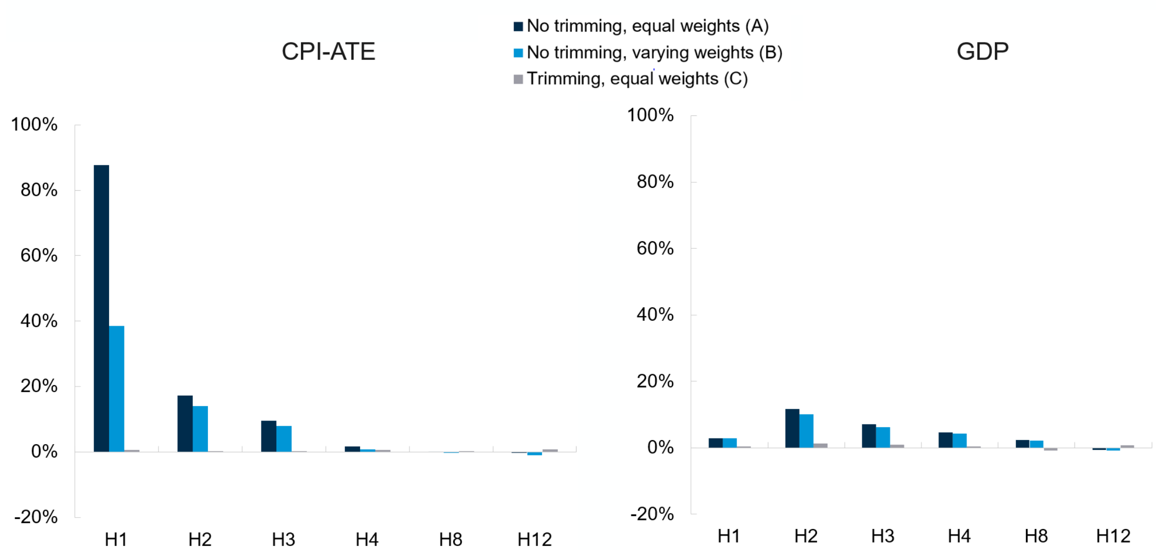


Figure 7: Relative RMSE to SMART. In percent. Positive values indicate worse forecasting-performance than SMART. By forecast horizon.

Figure 8 shows how RMSE changes subject to the discount factor for CPI-ATE (left-hand-side) and GDP (right-hand-side). For CPI-ATE, λ set to zero (the weights are purely based on the last available forecasting error) greatly increases the forecasting error for short-term horizons. This supports the idea that models with historically good forecasting performance will produce good forecasts in the future. Within the upper tier of discount factors the forecasting error does not change substantially. The results are similar for GDP, but with effect on forecasting performance.

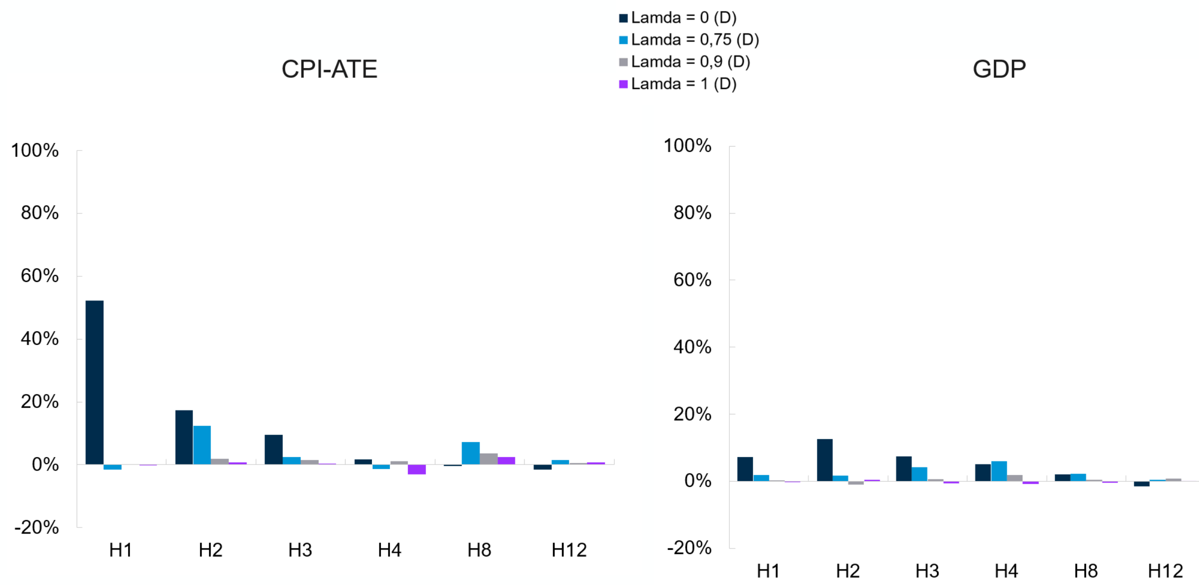


Figure 8: Relative RMSE to SMART. In percent. Positive values indicate worse forecasting-performance than SMART. By forecast horizon.

The SMART modelling system produces forecasts at all relevant context dates when new information is available, but a key consideration is deciding at which context dates to evaluate historical forecasting performance when combining the models. The current version of SMART is combined based on the forecasting performance on context dates close to historical MPR cut-off dates[5], at the end of each quarter. The main reason for selecting weights in SMART based on the forecasting performance at cut-off dates, is to calibrate the framework to produce the best possible forecasts close to the dates when the interest rate decisions are made. At this point, the available information set is more complete than when the policy process starts, roughly six weeks earlier. Figure 9 shows how the forecasting performance between the release date of quarterly national accounts (which is close to the start of the monetary policy process in Norges Bank) and the cut-off dates of MPR. Overall, it seems that the forecasting performance is good throughout the MPR process even though the weights are based on the cut-off dates.

[5]Cut-off dates are usually four business days prior to MPR-publications. Since CPI often is published after cut-off, but still incorporated in MPR-forecasts, SMART Inflation is evaluated at the context of publishing dates for MPR.

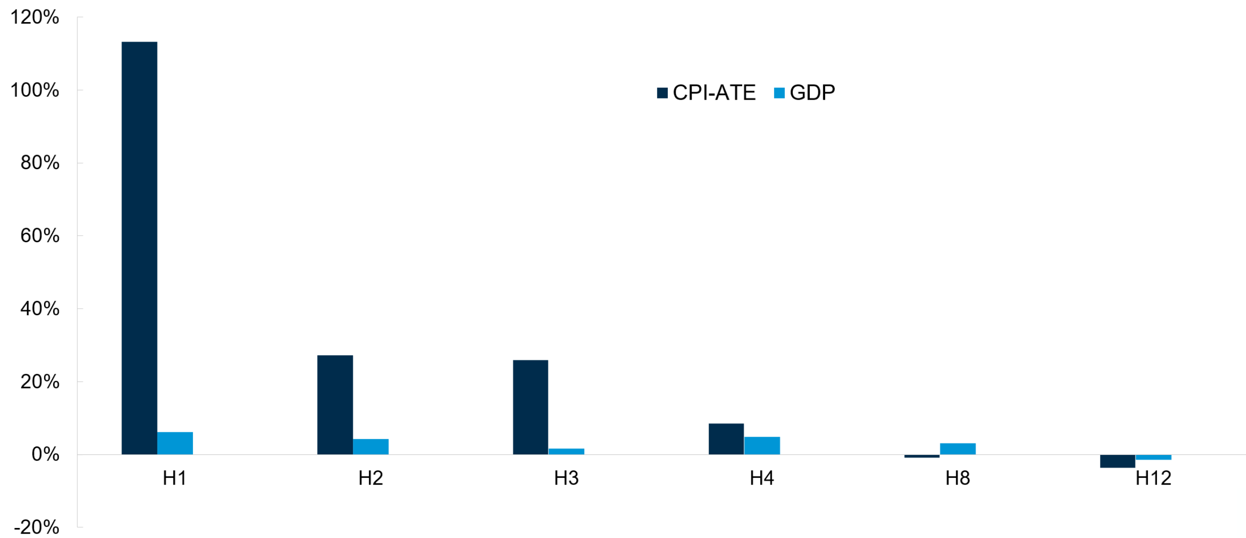


Figure 9: RMSE of SMART evaluated on QNR-dates relative to MPR-dates. In percent. Positive values indicate worse performance. By forecast horizon.

5 SMART Inflation

5.1 Overview of SMART Inflation

The current version of SMART Inflation is constructed to produce forecasts of Norwegian core inflation as measured by the CPI adjusted for tax changes and excluding energy products (CPI-ATE). SMART Inflation evaluates models based on their ability to forecast the four-quarter change in the CPI-ATE. To produce CPI forecasts for MPR, we add forecasts for energy prices based on energy futures prices to the CPI-ATE forecasts.

Following the introduction of inflation targeting in 2001, Statistics Norway started publishing CPI-ATE, with observations beginning in 1985. Norges Bank has calculated monthly CPI-ATE back to July 1979 for monetary policy use. The first vintage tag for CPI-ATE in SMART is June 2009[6].

SMART Inflation forecasts CPI-ATE decomposed by delivery sectors, domestic and imported inflation and by three-digit COICOP groups, 12 in all. By decomposing the inflation dynamics, we can identify its main drivers, which is crucial in both forecasting and in monetary policy analysis. Publishing forecasts decomposed by delivery sector also adds value to the narrative communication around the inflation outlook.

SMART Inflation now consists of 15 model ensembles that are trimmed and weighted according to principles from Section 4. The 15 ensembles include five model types: univariate models, small VAR-models, factor models, BVAR-models as well as single equation models (SEM) (Figure 10). All model types forecast total CPI-ATE and aggregated from the two delivery sectors. The univariate models also project price growth aggregated from the 12 consumption groups (see Appendix E.2). Univariate models, small VAR-models and factor models forecast inflation on a monthly frequency, also converted to quarterly forecasts (see Appendix E.3). In total, SMART Inflation currently consists of approximately 500 models. SMART Inflation uses of a broad set of macroeconomic and financial indicators. For an overview of statistics used in SMART Inflation, see Table 3.

[6] We constructed quasi-vintages prior to 2009.

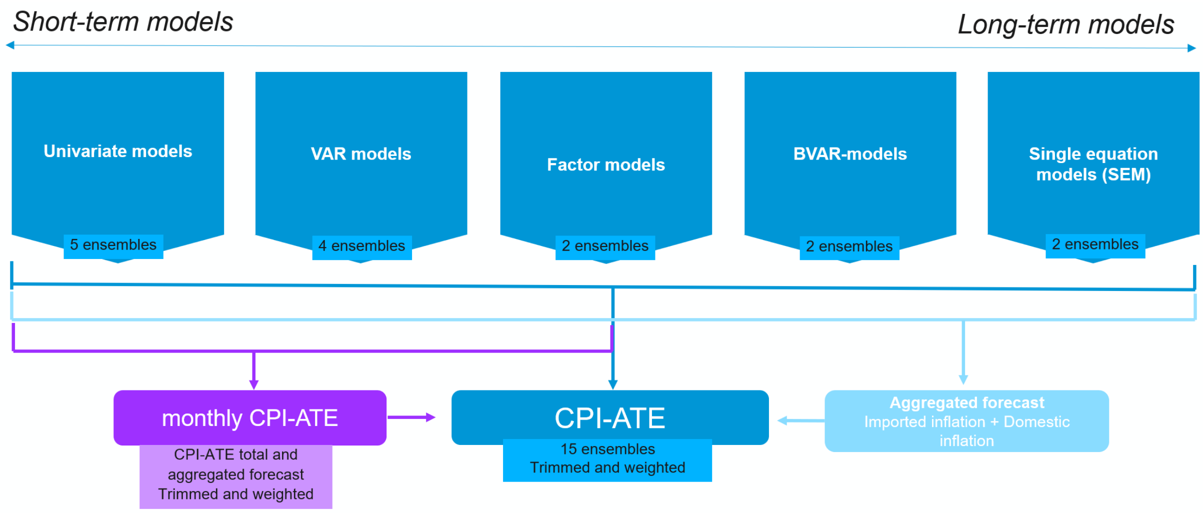


Figure 10: Overview of SMART Inflation

5.2 The models and ensembles in SMART Inflation

Univariate models

Univariate models (UNIVAR) are useful for capturing the intrinsic time series dynamics of inflation and will typically have decent forecasting properties in the short-run. In total, 200 univariate models across five model ensembles are included in SMART Inflation: CPI-ATE (quarterly and monthly), CPI-ATE aggregated from the delivery sectors (quarterly and monthly) and CPI-ATE aggregated from the 12 consumption groups (monthly). We vary the univariate models along several dimensions, for example lag structure and transformation. Some models are estimated with full sample while others are estimated with a rolling window of 10 years.

Small VAR-models

Small VAR-models are useful to capture simple, short-term dynamics between inflation and other key variables for forecasting. SMART Inflation includes over 100 bivariate VAR-models in four model ensembles that forecast CPI-ATE and CPI-ATE by two delivery sectors. The models are estimated at monthly and quarterly frequency. The VAR-ensembles include models with oil prices, exchange rates, producer prices, wage costs, profitability indicators, unemployment, capacity utilisation and PMI. The models are estimated with different model specifications, such as lag-length and data transformations. We handle unbalanced data, stemming from different publication lags of macro-data, by using suitable estimation and conditional forecasting techniques.

Factor models

Factor models (FACTOR) compliment univariate and small VAR models by capturing common dynamics across a large set of variables. SMART Inflation includes two model ensembles of factor models: one for the CPI-ATE and one for the delivery sectors. Some of the factor models use large sets of macroeconomic and financial data, including a mixed-frequency dynamic factor model with time varying parameters and stochastic volatility following Eraslan and Schröder (2022). We also have factor models for one- and two-digit COICOP groups. In total, 13 factor models are included in SMART Inflation.

Bayesian VAR-models

Bayesian VAR-models (BVAR) make up two model ensembles in SMART Inflation, one for CPI-ATE and one for the delivery sectors. The BVAR models are larger than the classic VAR-models and include both unconditional forecasts and forecasts conditioned on MPR forecasts for some of the variables.[7] In total, we include approximately 20 BVAR-models in SMART Inflation. The BVAR models capture larger dynamics across many variables and have useful longer term forecasting properties.

Single equation models (SEM)

The last two ensembles consist of estimated equations based on simple economic relationships such as the Phillips curve. SEM-models combine empirical relationships with simple economic theory and are typically good at medium to long-term forecasts. The group includes Autoregressive Distributed Lag models (ARDL) and Error Correction models (ECM). All models are estimated on a quarterly frequency, and conditioned on Norges Bank's MPR-forecasts. Like the other model ensembles, the SEM-models forecast CPI-ATE directly and aggregated by the two delivery sectors. These models are also useful as cross-checks as they condition on forecasts of the explanatory variables.

5.3 The SMART Inflation forecast

The combined SMART forecast follows the principles in Section 4 based on forecasts evaluation at the publishing dates of the MPR. SMART forecasts for inflation are available from 2003 Q1 (Figure 11).

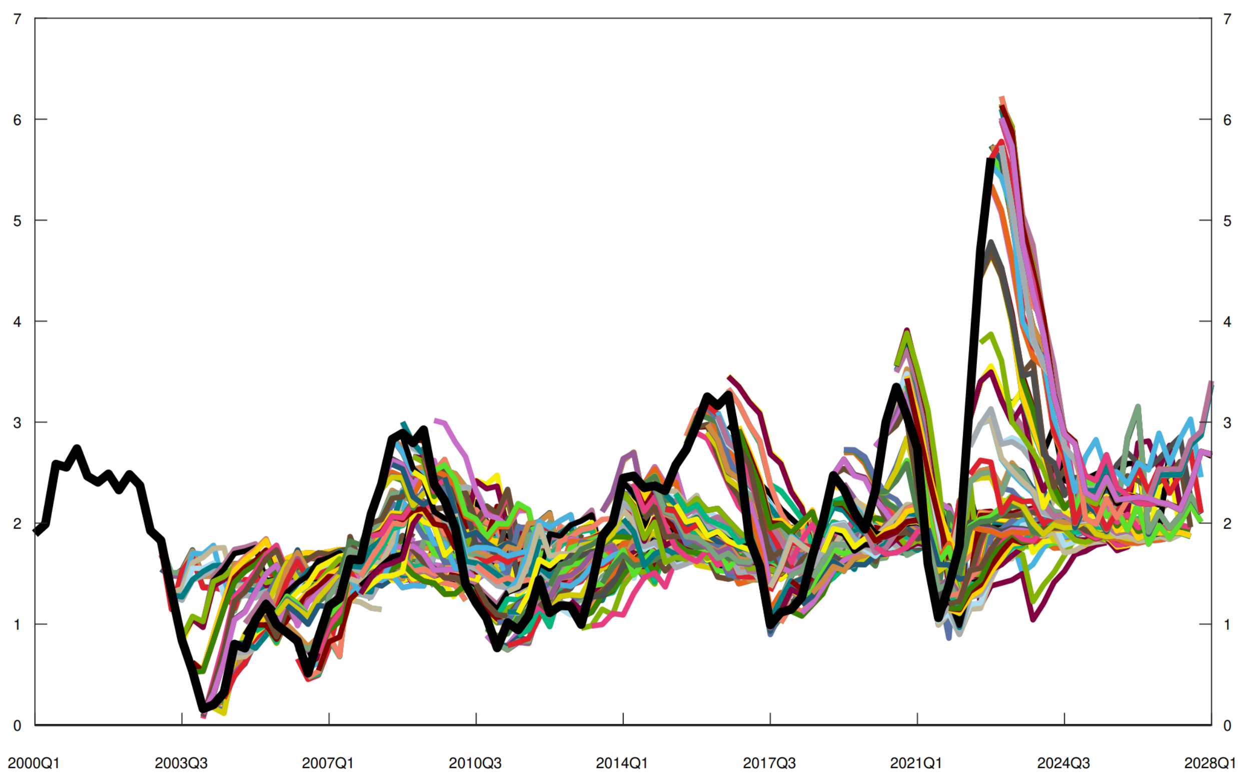


Figure 11: SMART Inflation. Four quarter change. 2003 Q1 - 2022Q4

[7] We use the prior of Giannone et al. (2015)

Figure 12 shows the performance by each individual ensemble compared to the final combined SMART forecast. Some ensembles perform slightly better at a single horizon, but SMART gives the best overall forecasts. Figure 13 shows which model ensembles that are given weights at different horizons. For the shorter horizons, univariate models, the factor models and the small VAR-models are given most weight. At longer horizons, more structural models, like the BVARs and the SEMs perform better. The results indicate that statistical and data driven models are better at short-term forecast, while the more structural models with some conditional information perform better in the longer-run.

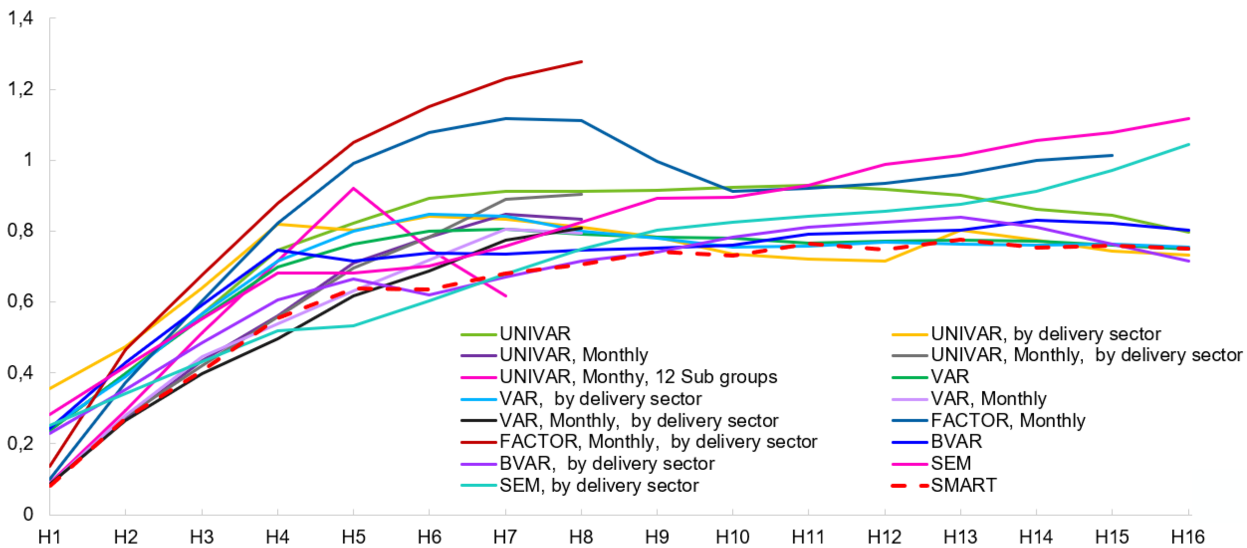


Figure 12: Forecasting performance of SMART Inflation and its model ensembles. Four quarter change. RMSE. By quarters. 2003 Q1 - 2019 Q4

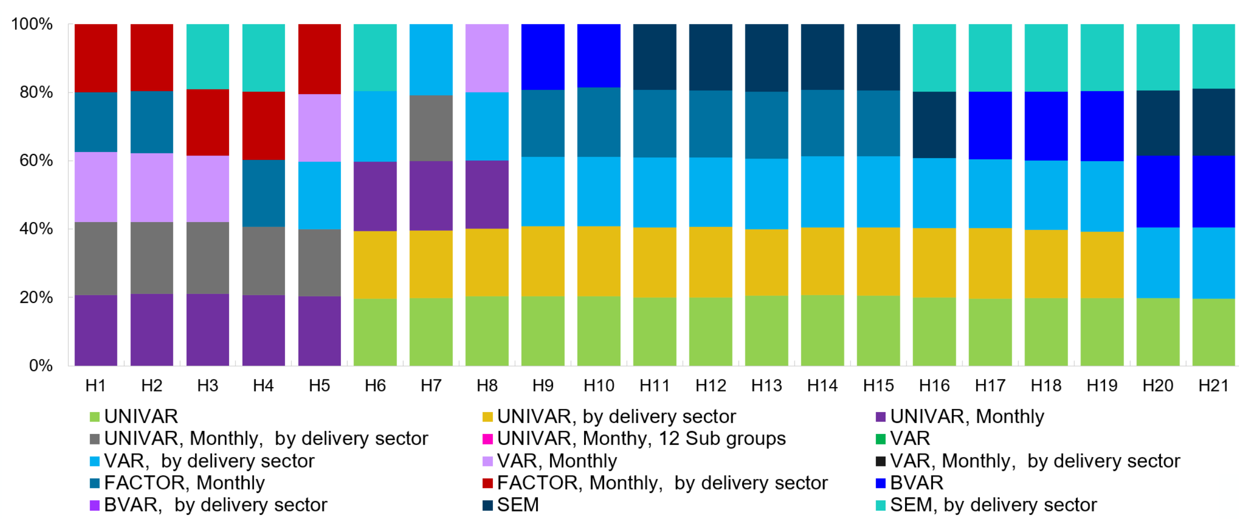


Figure 13: Weights to model ensembles in SMART Inflation. Based on DRMSE. By quarters. 2023 Q1

6 SMART GDP

6.1 Overview of SMART GDP

SMART GDP is calibrated to produce the best forecasts of the 12th release of seasonally adjusted, real GDP measured as the four-quarter growth rate. When measuring Norwegian GDP, it is common to exclude value added by the petroleum sector and from international shipping activities (see Fløttum et al. (2012)). Consistently, SMART GDP forecast the gross domestic product for mainland Norway. Quarterly observations of mainland GDP (hereinafter GDP) are available from 1978 Q1 in the quarterly national accounts (QNA) from Statistics Norway. Norges Bank has collected real-time observations of seasonally adjusted GDP since June 2000.

Monthly GDP has been published since September 2018, and is available from January 2016. We have estimated historical data beginning in 1990 (see Section 3.2). By including the monthly observations in the SMART system, we get a more updated forecast for the economic development.

SMART GDP is based on the best performing models from SAM, but other indicators and model types have been added. A broad set of macroeconomic and financial indicators are employed. For an overview of data used in SMART GDP, see Table 3. In contrast to SAM, the SMART forecasts for GDP also comprise of forecasts for sub-components of the gross product, as measured from the production side, as well as from the expenditure side. This allows for a consistency across components and is useful for the communication of the economic outlook.

The current version of SMART GDP includes 10 model ensembles that are trimmed and weighted to a single GDP-forecast based on the principles outlined in Section 4. The 10 ensembles include univariate models, small VAR-models, small direct forecast models, larger factor models and BVAR-models (Figure 14). Univariate models and small VAR-models are also used to forecast monthly GDP, converted to quarterly forecasts (see Appendix E.3). Additionally, we aggregate forecasts of eight demand components and 38 production sectors to forecasts of GDP (see Appendix E.2). Currently, SMART GDP consists of approximately 600 models.

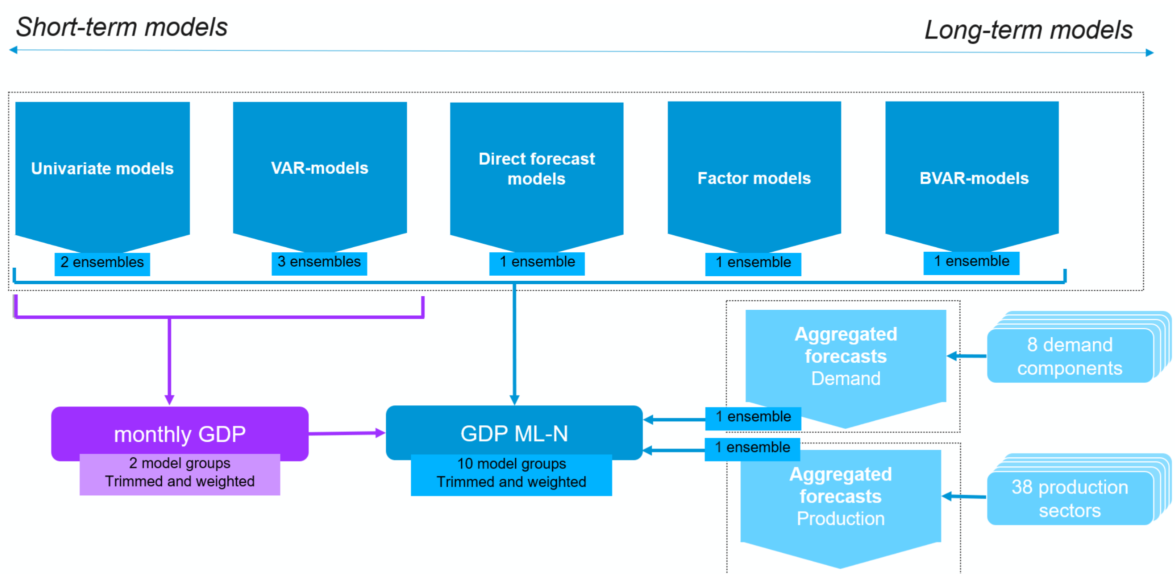


Figure 14: Overview of SMART GDP

Mainland GDP was substantially affected by the Covid pandemic. Between 2019 Q4 and 2020 Q2, GDP dropped by 8 percent, before recovering by 6 percent in 2020 Q3. The period had severe effects on parameter estimates. To avoid this, we either use Covid-dummies, remove observations from the estimation sample or model the large volatility directly, as suggested by Lenza and Primiceri (2020).[8] Models with stochastic volatility have shown the ability to handle the extreme observations during Covid. In models with stochastic volatility only the most extreme observations from 2020 Q2 are discarded. Further, when we evaluate the models, we discard forecasts both for and during the pandemic years.

6.2 A brief description of models and ensembles in SMART GDP

Univariate models

Similar to Smart Inflation, univariate models (UNIVAR) have proven useful for forecasting GDP. In total, around 200 univariate models, across two ensembles are included in SMART GDP: one at monthly and one at quarterly frequency. The model groups include a broad range of AR- and ARMA-models at different transformations, lag lengths and moving-average specifications.[9] We estimate these models on the full sample and with rolling windows of 20 to 40 quarters.

Small VAR-models

We capture short-term dynamics between GDP and one to two other key indicators in small VAR-models, organised into three model ensembles: one for soft data, such as survey indicators (VAR SOFT), one for hard data, such as macroeconomic statistics and financial market indicators (VAR HARD),[10] and one forecasting monthly GDP (VAR Monthly). The VAR-models are estimated for different model specifications. Lag length and data transformation are essentially chosen at discretion. In total, we include 140 small VAR-models with soft data from six different macro surveys, 160 small VAR-models with hard data from a variety of macro statistics and financial indicators, as well as around 40 monthly, small VAR-models. Similar to SMART Inflation, we utilise updated data by handling the unbalanced data problem.

Direct forecast models

To supplement small VAR-models, we included a separate model ensemble of 18 direct forecast models, or step-ahead models. Higher frequency data are bridged to estimate quarterly GDP. Only one model specification is included per indicator set.

Factor models

By exploiting information from large datasets, factor models can capture underlying tendencies. The current version of SMART GDP includes 12 dynamic factor models (FACTOR). The large monthly and quarterly factor models from SAM (Aastveit et al., 2011) are both included in SMART GDP, including approximately 80 indicators each. In addition, the factor model ensemble consists of a large mixed-frequency dynamic factor model with time varying parameters and stochastic volatility

[8]We have also tested for use of Covid indicators to adjust for variation in containment measures and infection rates, like the Covid-19 Stringency Index from the University of Oxford. Including such indicators in our models did not seem to work for Norway.

[9]Models with MA-processes are largely unstable but show some good forecasting performances at shorter horizons. We included a set of MA-models, but punish extreme forecasts as described in Section 4.

[10]Distinguishing between soft and hard data has been shown useful in nowcasting and short-term forecasting when facing data with publication lags (Bańbura and Rünstler, 2011).

(Eraslan and Schröder, 2022). Additional factor models are estimated on grouped selections of data, for example survey data, financial data, and demand components.

Bayesian VAR-models

The system includes several larger BVAR-models, grouped into one model ensemble. The BVARs are either unconditional or conditioned on MPR-forecasts. Some models are based on topical clusters of data, such as data for the household sector, the business sector, and the foreign sector. In total, we included five BVAR models in SMART GDP.[11] We will continue to further develop and improve the BVAR-ensemble in SMART GDP, to better capture longer-term dynamics across many variables.

GDP aggregated from the production side

GDP is measured from the production side, aggregating gross product in four main production sectors: Manufacturing and mining, production of other goods, private services and the public sector. The main production sectors can be decomposed into sub-sectors, defined by the Norwegian Standard Industrial Classification (SN2007), based on NACE Rev.2 (Statistics Norway, 2023). In Norges Bank, we forecast GDP across 38 production sectors.

To integrate the production sector approach in SMART GDP we have done some data adjustments. Gross production in commercial and non-commercial sectors are aggregated, and structural breaks are adjusted for. In the current version of SMART GDP, sector production is forecasted by using combinations of univariate models and small VAR-models. The sectors are also forecasted in a large BVAR-model with horseshoe prior (Carvalho et al., 2010).

Model forecasts are trimmed and weighted to a single sector forecast by using a one-step weighting scheme. The sector forecasts are summed to forecast GDP (PRODUCTION). This approach helps us capture sector specific trends and changes. Disaggregated GDP-forecasts have also been useful for storytelling in the forecasting process.

GDP aggregated from the demand side

GDP can also be aggregated from the demand side. The sum of forecasts of the demand components is included as a separate ensemble in SMART GDP. Norges Bank focus on seven demand components: private consumption, public demand, housing investment, investment in the petroleum sector, other business investment, mainland exports and mainland imports. Inventories and statistical deviations are defined as the difference between GDP and the sum of the demand components. Two aggregates from the demand side are included in SMART (DEMAND): one summarising forecasts for the seven demand components and one also including the inventories.

We use MPR and SMART forecasts for the demand components. Quarterly MPR-forecasts begins from 2010 for most of the demand components. For the period prior to 2010, simple forecasting models are used to extend the forecasting history for the demand components. In the future, the aim is that all demand components will have a separate SMART.

Consistency across expenditure sectors is crucial for the MPR forecasting process. By decomposing GDP into demand components, we can identify the main drivers of economic developments, which also adds value when communicating the economic outlook.

[11] Also these models are estimated using the GLP-prior of Giannone et al. (2015).

6.3 The SMART GDP forecast

The combined SMART GDP forecasts are based on the principles outlined in Section 4, and SMART forecasts for GDP are available starting in 2001 Q2 (Figure 15). SMART GDP reflects the forecasting performance of models at publication cut-off dates for MPR, towards the end of each quarter.

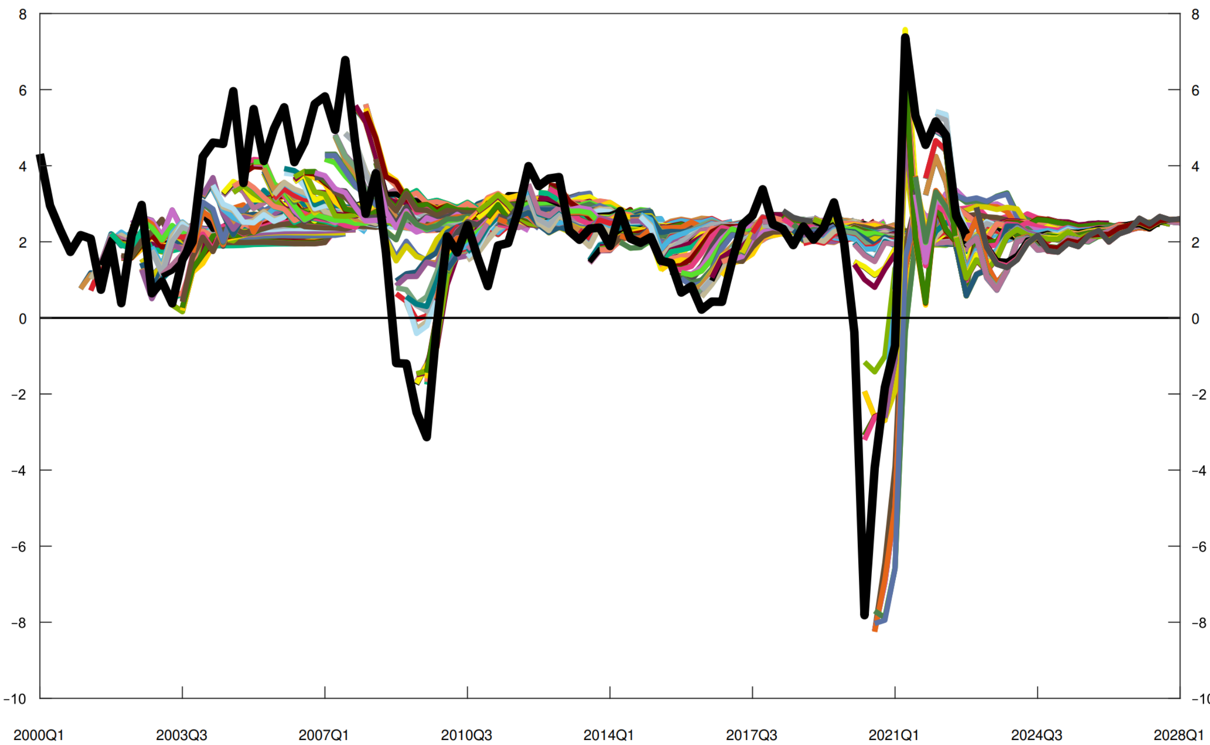


Figure 15: SMART GDP. Latest available vintage. Four quarter change. 2001 Q2 - 2019 Q4

On average, quarterly Norwegian GDP is revised around 4 percent between the first and the final release. SMART is evaluated against the 12th release of GDP.[12] By evaluating the GDP-models against the 12th release, we aim to forecast the most accurate measure of GDP.[13] For forecasts where the 12th release is not yet available, the forecasts are evaluated against the latest available release.

Figure 16 shows the forecasting performance by each individual ensemble compared to the final combined SMART forecast. Small VAR-models with survey indicators and factor models seem to be the best performing ensembles at the shorter horizons. Consequently, together with the direct forecast-models, their forecasts are given most weight in the short-run (Figure 17). For medium- and longer-term forecasts it is primarily larger multivariate models and aggregated forecasts from the demand sectors that receive weight. Similar to SMART Inflation, the results indicate that statistical and data driven models are better at short-term forecasting, while larger, more structural models with some conditional information perform better in the longer run. The combined SMART GDP

[12]Final numbers for GDP are normally published in August each year, so that Q1-observations are final on the 10th release, while Q4 observations are final on the seventh. Occasionally, the final release has been postponed to the November-publication of QNA. For forecasts where the 12th release is not yet available the forecasts are evaluated against the latest available release.

[13]The SMART forecasts also forecast the first release of GDP reasonably well (see C).

forecast performs well compared to the individual model ensembles; however, the factor ensemble performs slightly better. We still believe that combining ensembles is the best strategy for forecasting and expect SMART GDP to further improve as we continue to develop new models and add more ensembles.

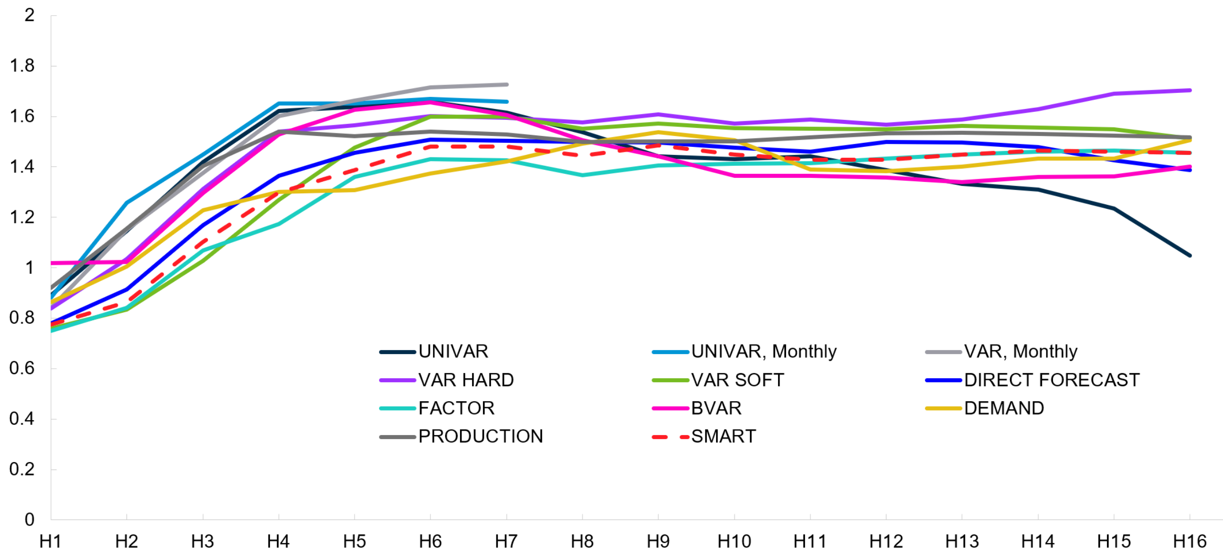


Figure 16: Forecasting performance of SMART GDP and its model ensembles. Four quarter change. RMSE. By quarters. 2002 Q1 - 2019 Q4

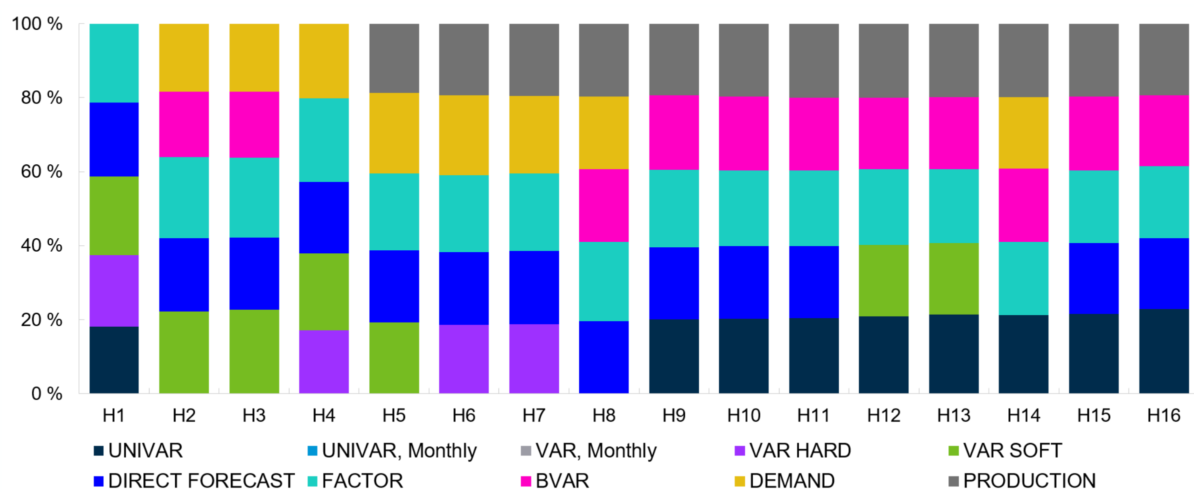


Figure 17: Weights to model ensembles in SMART GDP. Based on DRMSE. By quarters. 2023 Q1

7 Evaluation

In order to evaluate the performance of SMART, we preform an historical out-of-sample forecasting exercise of SMART against certain benchmark models for the four-quarter change in CPI-ATE and GDP in real-time, summarized in table 1. GDP is evaluated against the 12th release. The main focus of SMART is short-term forecasting, but the system is also used for cross-checks of medium-term monetary policy forecasts. The forecast evaluation is reported for horizons 1-4, 8, 12 and 16. We evaluate model forecasts given at dates close to historical MPR-publications.[14]

The historical forecast performance of SMART is evaluated against three benchmarks with different forecasting properties.

1. A naïve AR-model
2. Historical forecasts from our previous forecasting system, SAM
3. Norges Bank's historical MPR forecasts

forecasts	Period	Horizons	Forecast type
AR	2001 Q2 - 2019 Q4	16	Unconditional, point forecast
SAM	2009 Q1 - 2019 Q4	5	Unconditional, density forecast
MPR	2010 Q1 - 2019 Q4	13-16	Structural model/judgment, point forecast
SMART	2001 Q2 - 2019 Q4	16	Conditional, point forecast

Table 1: SMART forecasts and relevant benchmarks

The AR benchmark models for the CPI-ATE and GDP is an AR(2) and AR(4), respectively. These are the AR-processes that best describe the two time series. SMART performs significantly better than the AR benchmark at the first two horizons (Table 2). The difference in forecasting performance decreases as the horizon increases.

SMART is also evaluated against our previous forecasting system, SAM, and historical MPR forecasts[15]. As shown in Table 2, SMART performs equally as good (or better) than both SAM and MPR forecasts, but the difference is not statistically significant for most horizons.

[14] Prior to 2013, MPR was only published three times a year. We therefore simulated a quarterly calendar matching the timing of MPR publications prior to 2013. The benchmark against MPR is only evaluated at dates where actual MPR forecasts existed

[15] For these benchmarks the forecast history is shorter, so the comparison is based on a shorter sample.

	SMART Inflation			*SMART GDP*		
	SAM	MPR	AR(2)	SAM	MPR	AR(4)
Period	09Q1-19Q4	10Q1-19Q4	03Q1-19Q4	09Q1-19Q4	10Q1-19Q4	02Q1-19Q4
H1	-0.01**	0.00	-0.2***	-0.02	0.00	-0.08**
H2	-0.03	-0.02	-0.16***	-0.08	-0.04	-0.16*
H3	-0.08	-0.02	-0.2	-0.16	-0.02	-0.17
H4	-0.08	-0.00	-0.18	-0.15	-0.02	-0.17
H8		-0.04	-0.05**		-0.15	-0.04
H12		-0.01	0.01*		-0.25	-0.04
H16			0.05			0.02

*RMSE from SMART relative to RMSE from benchmark model. The difference is tested by Diabold Modigliani at * 5 percent significance, ** 1 percent, *** 0.1 percent.*

Table 2: Relative performance of SMART

8 Summary

Overall, we are confident that SMART will be a useful platform for forecasting macroeconomic variables. Having access to real-time data and models enables us to update forecasts in a timely manner when new data are available. SMART is already an integrated part of Norges Bank's forecasting process and SMART forecasts are regularly presented to Norges Bank's Monetary Policy and Financial Stability Committee. Short-term forecasts were published for the first time in MPR 1/23 (Figure 18).

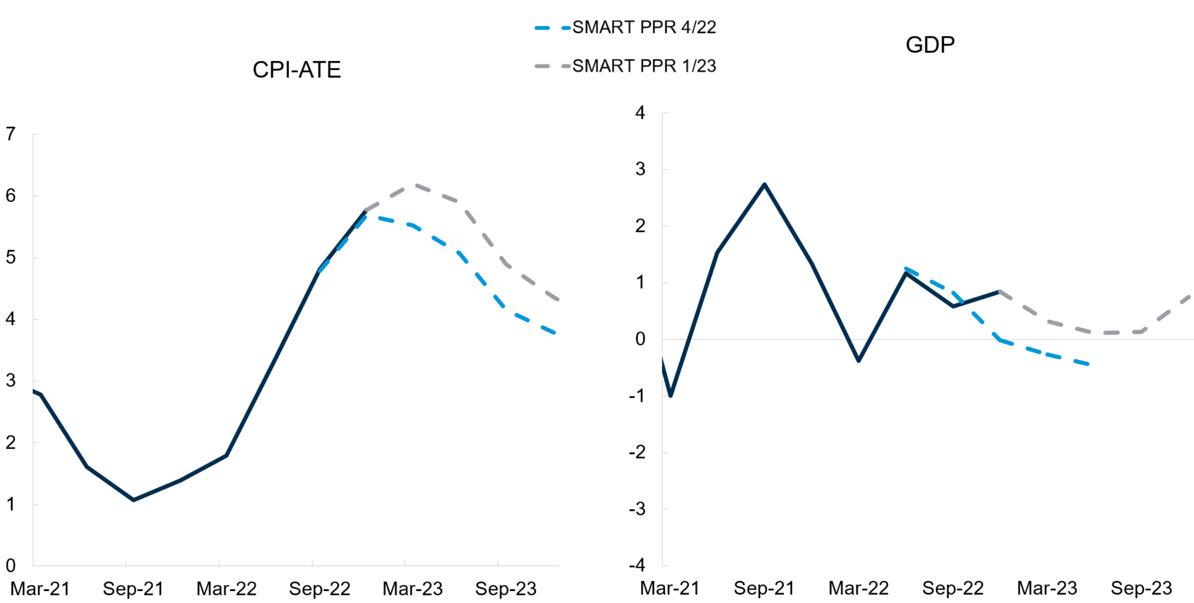


Figure 18: SMART forecast for inflation (four-quarter growth) and GDP (quarterly growth).

The SMART platform gives large possibilities in terms of model forecast analysis. For example, SMART is well suited to assess and illustrate model uncertainty, and Figure 19 illustrates how the SMART forecasts for 2023 Q1 and 2024 Q1 has changed over time. This is useful for understanding how new data releases change the economic outlook.

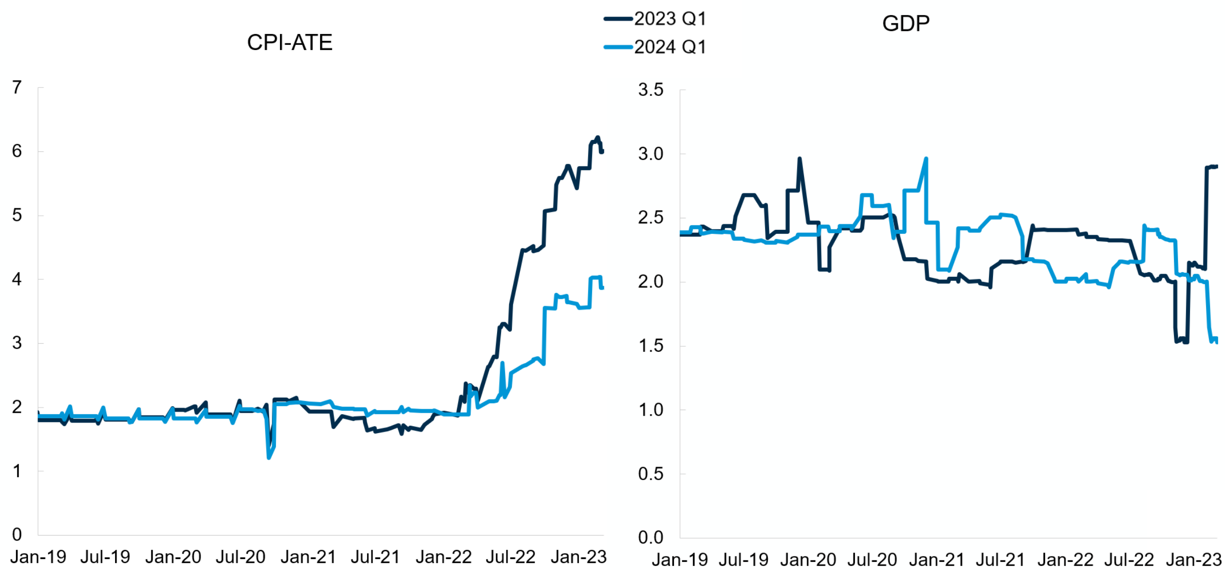


Figure 19: Revisions in SMART forecasts. CPI-ATE and GDP for 2023 Q1 and 2024 Q1.Four-quarter change.

This paper documents the first version of SMART Inflation and SMART GDP. We will continue to work on improving the systems for GDP and inflation and extend SMART to include all the main macroeconomic variables we forecast at Norges Bank as part of the monetary policy process. Furthermore, we are working on developing SMART to allow for scenario analysis and state-dependent weighting schemes. Our aim is also to include new data sources, such as debit and credit card transaction data and textual data from news articles. The aim is for SMART to be an agile and flexible tool that is useful Norges Bank staff when producing forecasts and analysing the macroeconomic outlook.

References

Aastveit, K. A., Fastbø, T. M., Granziera, E., Paulsen, K. S., and Torstensen, K. N. (2020). Nowcasting Norwegian household consumption with debit card transaction data. Working Paper 17, Norges Bank.

Aastveit, K. a., Gerdrup, K., and Jore, A. S. (2011). Short-term forecasting of GDP and inflation in real-time: Norges bank's system for averaging models. Staff memo 9, Norges Bank.

Aastveit, K. A., Gerdrup, K. R., Jore, A. S., and Thorsrud, L. A. (2014). Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics*, 32(1):48–68.

Bates, J. M. and Granger, C. W. J. (1969). The Combination of Forecasts. *OR*, 20(4):451–468.

Bańbura, M. and Rünstler, G. (2011). A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *International Journal of Forecasting*, 27(2):333–346.

Bowe, F., Gerdrup, K. R., Maffei-Faccioli, N., and Olsen, H. (2023). A high-frequency financial conditions index for norway. Staff memo 1, Norges Bank.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Clark, T. and McCracken, M. (2013). Chapter 20 - advances in forecast evaluation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 1107–1201. Elsevier.

Croushore, D. and Stark, T. (2003). A real-time data set for macroeconomists: Does the data vintage matter? *The Review of Economics and Statistics*, 85(3):605–617.

Diebold, F. X. and Shin, M. (2019). Machine learning for regularized survey forecast combination: Partially-egalitarian LASSO and its derivatives. *International Journal of Forecasting*, 35(4):1679–1691.

Elliott, G., Granger, C., and Timmermann, A., editors (2006). *Handbook of Economic Forecasting*, volume 1 of *Handbook of Economic Forecasting*. Elsevier.

Eraslan, S. and Schröder, M. (2022). Nowcasting GDP with a pool of factor models and a fast estimation algorithm. *International Journal of Forecasting*.

Fløttum, E. J., Halvorsen, T., Simpson, L. H., and Skoglund, T. (2012). History of National Accounts in Norway - From free research to statistics regulated by law. *SSB Social and Economic Studies*, 113.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics*, 97(2):436–451.

Jose, V. R. R. and Winkler, R. L. (2008). Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24(1):163–169.

Koop, G. and Korobilis, D. (2012). Forecasting Inflation Using Dynamic Model Averaging. *International Economic Review*, 53(3):867–886.

Kravik, E. and Mimir, Y. (2019). Navigating with NEMO. Staff memo 6, Norges Bank.

Lenza, M. and Primiceri, G. E. (2020). How to estimate a var after march 2020. Working Paper 27771, National Bureau of Economic Research.

Norges Bank (2020). Evaluation of Norges Bank's forecasts for 2018 and 2019. Norges Bank Papers 5, Norges Bank.

Norges Bank (2022). Monetary Policy Report with financial stability assessment 2/2022. https://www.norges-bank.no/en/news-events/news-publications/Reports/Monetary-Policy-Report-with-financial-stability-assessment/2022/mpr-22022/.

Paulsen, K. S. and Loneland, A. (2021). *NB Toolbox: Reference manual*. https://github.com/Coksp1/NBTOOLBOX.

Rossi, B. (2013). Advances in Forecasting under Instability. In Elliott, G., Granger, C., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2 of *Handbook of Economic Forecasting*, pages 1203–1324. Elsevier.

Schorfheide, F. and Song, D. (2015). Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, 33(3):366–380.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review*, 97(3):586–606.

Stark, T. and Croushore, D. (2002). Forecasting with a real-time data set for macroeconomists. *Journal of Macroeconomics*, 24(4):507–531.

Statistics Norway (2023). Classification of standard industrial classification. https://www.ssb.no/en/klass/klassifikasjoner/6. Accessed: 2023-02-07.

Thorsrud, L. A. (2018). Words are the new numbers: A newsy coincident index of business cycles. *Journal of Business & Economic Statistics*, 38(2):393–409.

A Data in the SMART database

This table summarises the statistics included in the SMART database, and where the statistics are used.

Data in SMART				
Data	Source	Frequency	SMART Inflation	SMART GDP
Domestic variables				
National accounts	SSB	Q/M	X	X
CPI	SSB	M	X	X
Capacity utilisation	NB	Q	X	X
Registred labour market data	NAV	M	X	X
Labour Force Survey	SSB	M/Q		X
Employments and earnings	SSB	M		X
Industrial production	SSB	M	X	X
PMI	DNB Markets, NIMA	M		X
Producer price index	SSB	M	X	
Accommodation	SSB	M	X	X
External trade	SSB	M	X	X
Index of wholesale and retail sales	SSB	M	X	X
Index of household consumption of goods	SSB	M	X	X
Housing prices	Eiendomsverdi, Finn.no, Real Estate Norway	Q		X
Housing starts	SSB	M	X	X
Construction cost index	SSB	M		
Credit indicators	SSB	M	X	X
Monetary aggregates	SSB	M	X	X
Non-financial sector accounts	SSB	Q/A		
Investment statistics	SSB	Q		
Vehicle registration	SSB/OFV	Q		
Air passengers	Avinor	W		
Tax	Minestry of Finance	A		
Interest rates (households and NFCs)	SSB/Finansportalen	D/M		X
Norges Bank's output gap	NB	Q	X	X
IPK	NB	Q	X	
Norges Bank's Regional network	NB	Q	X	X
Norges Bank's Expectations Survey	NB	Q	X	X
Business tendency survey (KBAR)	SSB	Q	X	X
KANTAR consumer confidence	KANTAR/Finans Norge	Q	X	X
NHO membership survey	NHO	Q		X
Norges Bank's Survey of Bank Lending	NB	Q		
Financial market variables				
Exchange rates	NB/Datastream	D	X	X
Interest rates	NB/Datastream	D	X	X
Stock market indexes	Datastream	D	X	X
CDS-prices	Bloomberg	D	X	X
Credit premiums	Nordic Bond Pricing	W		X
Foreign variables				
PMI	Datastream	M		X
Consumer and business confidence surveys	Datastream/OECD	Q		X
QNA	Datastream	Q		X
CPI	Datastream	Q		
Unemployment	Datastream	Q		
Production index for construction	Datastream	Q		
Industrial production	Datastream	Q		
Retail sales	Datastream	Q		
Vehicle registration	Datastream	Q		
Housing prices	Datastream	M		
Energy prices	Datastream/Nordpool	D	X	X
Commodity prices	Datastream	D	X	X

IPK: *Prices for imported consumer goods.*
Frequencies: *Q: quarterly, M: monthly, D: daily (business)*
Sources: *NB: Norges Bank. SSB: Statistics Norway. NAV: Norwegian Labour and Welfare Administration. NIMA: Norwegian Association of Purchasing and Logistics. Datastream: Refinitiv Datastream.*

B Estimating synthetic monthly GDP-history

Statistics Norway publishes monthly real GDP observations starting in January 2016. We estimate monthly GDP from January 1990 to December 2015 using a Bayesian Mixed Frequency-VAR (MF-VAR) approach as in Schorfheide and Song (2015) where we observe GDP on both monthly and quarterly frequencies and include different sets of indicators that may describe the monthly variation of GDP. We use the following state space model

$$J_t Y_t = H_t X_t + v_t, \tag{3}$$

$$v_t \sim N(0, R), \tag{4}$$

$$X_t = A X_{t-1} + u_t, \tag{5}$$

$$u_t \sim N(0, \Sigma). \tag{6}$$

where Y_t is the observed variables of the model with size M x 1, X_t is the unobserved state vector with size S x 1, v_t is the vector with measurement errors with size M x 1, R is the measurement error covariance matrix with size M x M, u_t is a vector with the reduced form residuals of the model with size S x 1. A has size S x S and Σ has size M x M, and are parameters to be estimated. The matrices J_t and H_t are time varying due to fact that we are dealing with missing observations. In the case we have no missing observations the matrix J_t will be the identity matrix with size M x M, while the H_t will be a matrix with size M x S. We come back to how we find H_t. When we have missing observations in the vector Y_t, we remove the corresponding rows of J_t and H_t, i.e. we only let the non-missing values in Y_t, inform the state vector X_t in period t.

In our case the measurement equation takes the form

$$J_t \overbrace{\begin{bmatrix} GDP_t^Q \\ GDP_t^M \\ IND_t^M \end{bmatrix}}^{Y_t} = \overbrace{\begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} & 0 & 1 & 0 & \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}^{H_t} \overbrace{\begin{bmatrix} GDP_t^M \\ IND_t^M \\ GDP_{t-1}^M \\ IND_{t-1}^M \\ GDP_{t-2}^M \\ IND_{t-2}^M \\ GDP_{t-3}^M \\ IND_{t-3}^M \\ GDP_{t-4}^M \\ IND_{t-4}^M \end{bmatrix}}^{X_t} + \overbrace{\begin{bmatrix} v_t^Q \\ v_t^M \\ v_t^{IM} \end{bmatrix}}^{v_t}. \tag{7}$$

where GDP_t^Q is quarterly real GDP, GDP_t^M is monthly real GDP, and IND_t^M is a vector of monthly indicators. We are using a dogmatic prior on R

$$R_i = \sigma_i \cdot R_i^{scale}, \tag{8}$$

We set $R_Q^{scale} = 0$, $R_M^{scale} = \frac{1}{100}$ and $R_I^{scale} = 0$, i.e. we allow for measurement error in the monthly real GDP series, but not in the others. σ_M is the estimated variance of monthly real GDP from January 2016 to December 2022 excluding observations of the year 2020.

We set up 500 models using a set of 12 monthly indicators; Industrial production, household consumption of goods, unemployment, the exchange rate, the oil price, a FCI for Norway, housing starts, house prices, accommodation and three indicators for credit to households and NFCs. The models uses five lags. We estimate these models using the prior of Giannone et al. (2015) and cross-validate them by removing one year of data on the monthly GDP, i.e. the years 2016-2021, but where we ignore the year 2020 due to the Covid-19 pandemic. We select the 20 best models among the 500 based on calculated RMSEs from the cross-validating. We weight these 20 models using inverse RMSE weights. As this weighted series does not add up to quarterly GDP, we estimate a final MF-VAR model using the weighted series as the only indicator. The output from this model is our estimate of monthly real GDP from January 1990 to December 2015 (Figure 20).

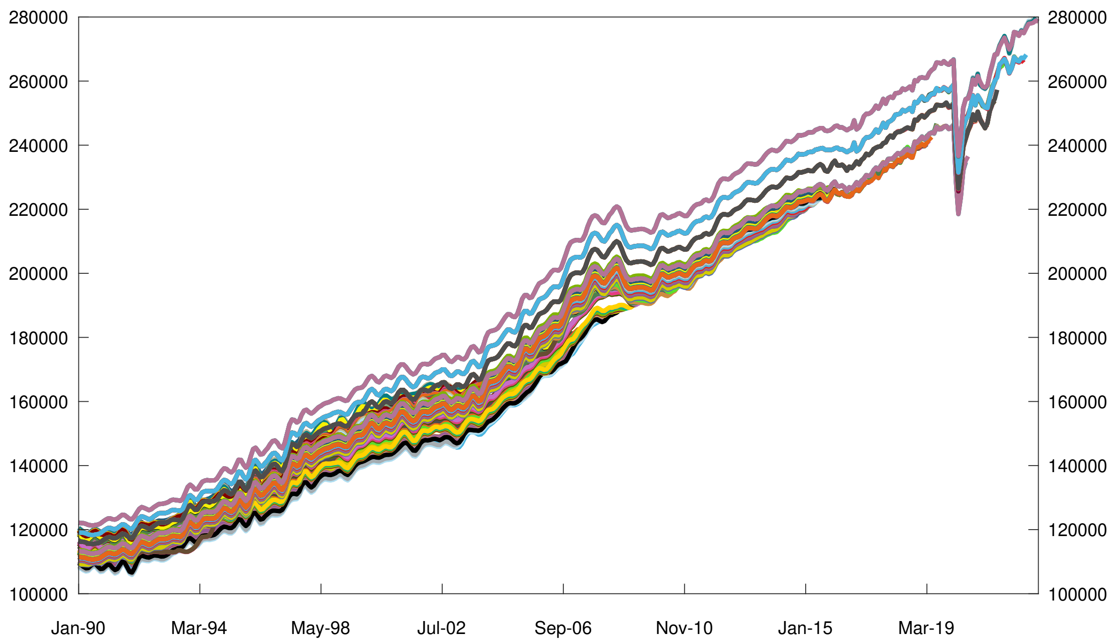


Figure 20: Estimated real-time vintages of monthly real GDP for mainland Norway. Seasonally adjusted. In billions of NOK. January 1990-January 2023. Publication dates between 2003-02-14 to 2023-03-16. All publication dates that do not match quarterly real GDP are estimated based on the publication lag of the first vintage of monthly real GDP published by Statistics Norway, which is 2017-11-14. This vintage ends in September 2017, i.e. with about 2 months publication lag. Before 2006-12-05 the publication lag is extended with one month.

C Consequence of targeting the 12th release of GDP instead of the 1st

SMART GDP is evaluated to project GDP data at its 12th release. Assuming that the final release is the the most accurate measure of the level of GDP, forecasts of the 12th release will over time result in the best calibration of monetary policy. However, a key element of MPR policy processes is to evaluate new observations, the 1st release, against monetary policy forecasts. Evaluated at its first release, SMART actually performs better compared with AR-forecasts at the same release (see chart 21).[16]

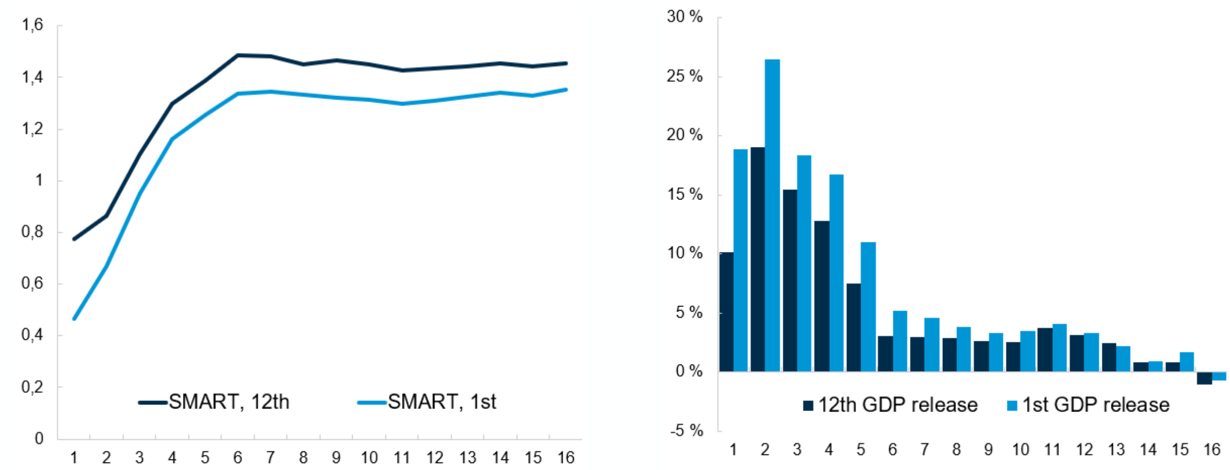


Figure 21: Left: RMSE for 1st and 12th release. Right: Relative RMSE of SMART vs. AR(4).In percent. Positive values indicate worse performance than SMART. By forecast horizon. Evaluated during the quarter.Four quarter change. 2010 Q2-2019 Q4

D Calculations in SMART

Let $S(c)$ be the number of observations at context c. Given a set of time series $X_{S(c)}$ we define a calculation as

$$y_{S(c)} = H(X_{S(c)}), \tag{9}$$

where H is a mapping from $H : \mathbb{R}^{N_{S(c)} \times M_{S(c)}} \longrightarrow \mathbb{R}^{N_{S(c)}}$, with $N_{S(c)}$ as the number of observations and $M_{S(c)}$ as the number of time series of $X_{S(c)}$. That means that the mapping H calculates a new variable y at each context c, using data on a set of other times-series at the same context c. The set of contexts will depend on which time series are used to calculate y.

E Forecasting in SMART

In this section we give a brief introduction to combination and aggregation of forecasts in SMART as well as how forecasts of models are converted from one frequency to another. We end the section by introducing a calculator and briefly highlight how they differ from models in SMART.

E.1 Combining Forecasts

Given as set of models, let us define the forecast of model m at a context c and time $s(c)$ at horizon h by $f_{s(c),h,m}$, and the actual data at time r as Y_r. The forecast error at the context c at horizon h is then given by $\eta_{s(c),h,m} = f_{s(c),h,m} - Y_{s(c)+h}$, if $Y_{s(c)+h}$ exists, otherwise $\eta_{s(c),h,m} = NaN$. To score the real-time forecasts a metric must be chosen. We use discounted root mean squared error(DRMSE) defined as

[16]This test is not relevant for SMART Inflation, as CPI-ATE is not revised.

$$score_{h,C,m} = \sqrt{\frac{\sum_{c \in C} \lambda^{N_c - n_c} \eta^2_{s(c),h,m}}{\sum_{c \in C} \lambda^{N_c - n_c}}}, \tag{10}$$

where C is the set of context dates from the given model matching a selected calendar. N_C is the number of dates matched by the calendar, n_c is the index of the context c in C and λ is the discount factor.

If a model m gives explosive forecasts at a context c we set $f_{s(c),h,m} = NaN$[17][18], we define a explosive forecast as forecast that is lower than $l^-_{s(c),h}$ or higher than $l^+_{s(c),h}$. These limits are defined

$$l^{\pm}_{s(c),h} = f_{s(c),h,me} \pm \sigma_{\tau,s(c)} K, \tag{11}$$

where $f_{s(c),h,me}$ is the median forecast among all models producing forecast at context c, $\sigma_{\tau,s(c)}$ is calculated as the standard deviation of the last τ observations of the context c of the forecasted variable, and K is a parameter chosen by the user.

If a model does not produce a forecast at context c, we punish that model with

$$(\eta_{s(c),h,m})^2 = (\eta_{s(c),h,a})^2 + \sigma^P_{s(c)} W, \tag{12}$$

where $(\eta_{s(c),h,a})^2$ is the mean of the squared forecast error of all models providing a valid forecast, and $\sigma^P_{s(c)}$ is the corresponding standard deviation. W is a parameter chosen by the user.

We combine forecasts, using the DRMSE scores. The combined forecast is just a weighted mean of the models forecasts, i.e.

$$CF_{s(c),h} = \sum_{m=1}^{M} w_{h,C(c),m} \times f_{s(c),h,m}, \tag{13}$$

where $CF_{s(c),h}$ is the combined forecast, $f_{s(c),h}$ is the individual model forecasts and $w_{h,C(c),m}$ is the weights of each model m, which are calculated using real-time forecast scores up until the given context c, abbreviate as $C(c)$, and is constructed as follows

$$w_{h,C(c),m} = 1_{\{m \in Z(M)_{h,C(c)}\}} \times \frac{\frac{1}{score_{h,C(c),m}}}{\sum_{\mu \in Z(M)_{h,C(c)}} \frac{1}{score_{h,C(c),\mu}}}, \tag{14}$$

where 1 and $Z(M)_{h,C(c)}$ are defined as in section 4, but now we take into account that it depends on the given context c.

[17] NaN is short for *Not a Number*. It is the standard in most programming languages for representing a floating-point operations that have some input parameters that cause the operations to produce undefined results. For example, 0/0 or ∞/∞. It is also used for representing missing values.

[18] We handle operations involving NaN values different depending on context. In some cases operations involving one or more values of NaNs will result in NaN, while in other cases the operations will result in a (real) number, i.e., the NaN value(s) are treated either as 0's or not existing.

E.2 Aggregating Forecasts

In the case we have forecasts on a set, V, of disaggregated variables $f_{s(c),h,v}$, with $v \in V$. We may aggregate these forecasts using the known weights $w_{s(c),h,v}$ as a weighted mean of the forecasts of the disaggregated variables, i.e.

$$AF_{s(c),h} = \sum_{v=1}^{V} w_{s(c),h,v} \times f_{s(c),h,v}, \tag{15}$$

where the weights at horizon h are predicted using the random walk model (constant forecast):

$$w_{s(c),h,v} = w_{s(c),v}. \tag{16}$$

E.3 Converting Forecasts

If we have forecasts on high frequency $f_{s^{hf}(c),h}$, and want to convert these to low frequency forecasts $f_{s(c),h}$ we can define a mapping function

$$f_{s(c),h} = G(\{f_{s^{hf}(c),h^{hf}}\}_{h^{hf} \in P(h)}), \tag{17}$$

where $s^{hf}(c)$ is the corresponding observed date of the last high frequency observation in the low frequency period $s(c)$, h and h^{hf} are the horizons of the low frequency and the high frequency forecasts, respectively, and $P(h)$ is the index set of all high frequency horizons in the low frequency horizon h. For example, when $s(c)$ is 2022Q4 and $s^{hf}(c)$ is 2022M12, i.e., quarterly and monthly frequencies, respectively, then $P(1)$ is the set $\{1, 2, 3\}$, $P(2)$ is $\{4, 5, 6\}$ and so on. Similarly, for annually and monthly frequencies, i.e., if $s(c)$ is 2022 and $s^{hf}(c)$ is 2022M12, then $P(1)$ is the set $\{1, 2, \ldots, 12\}$, $P(2)$ is $\{13, 14, \ldots, 24, \}$ and so on. The mapping function G can be taken as the sum, average or as the last observation of the high frequency forecasts.