No. 09 | 2011

Staff Memo

Short-term forecasting of GDP and inflation in real-time: Norges Bank's system for averaging models

Knut Are Aastveit, Karsten R. Gerdrup and Anne Sofie Jore, Monetary Policy

NB NORGES BANK

ISSN 1504-2596 (online only)

ISBN 978-82-7553-617-2 (online only)

Short-term forecasting of GDP and Inflation in Real-Time: Norges Bank's system for averaging models*

Knut Are Aastveit† Karsten R. Gerdrup‡ Anne Sofie Jore§

August 16, 2011

Abstract

In this paper we describe Norges Bank's system for averaging models (SAM) which produces model-based density forecasts for Norwegian Mainland GDP and inflation. We combine the forecasts from three main types of models typically used at central banks: Vector autoregressive models, leading indicator models and factor models. By combining models we hedge against uncertain instabilities. We update SAM several times during the quarter to highlight the importance of new data releases, and we show how the performance of SAM improves steadily as new information arrives. The framework is robust with regard to alternative vintages of data to evaluate against. We show that our chosen weighting scheme is superior or on a par with some common alternative weighting schemes, and, finally, that a strategy of trying to pick the best model, ex ante, is inferior to model combination.

JEL-codes: C32, C52, C53, E37, E52.

Keywords: Density combination; Forecast densities; Forecast evaluation; Monetary policy; Nowcasting; Real-time data

*The views expressed in this paper are those of the authors and should not be attributed to Norges Bank.

†Norges Bank, Economics Department, Knut-Are.Aastveit@norges-bank.no

‡*Corresponding author*: Norges Bank, Economics Department, Karsten.Gerdrup@norges-bank.no

§Norges Bank, Economics Department, Anne-Sofie.Jore@norges-bank.no

1 Introduction

Policy decisions in real-time are based on assessments of the recent past and current economic condition under a high degree of uncertainty. Short-term forecasts at central banks are typically formed by sector experts' views on economic developments in conjunction with formal models. In this process, forecasters have to take into account that statistics are released with a long delay, are subsequently revised and are available at different frequencies. In addition, the data generating process is typically unknown and likely to change over time.

Having a good understanding of current economic conditions is important, because it provides policy makers with a starting point for medium- to long-term forecasts and policy analysis. Norges Bank's criteria for what constitutes a good model of the economy depends, broadly speaking, on the time horizon of forecasts and analysis, see Gerdrup and Nicolaisen (2011). In the short-run, empirical fit and out-of-sample performance are the primary concerns. In the medium- to long-term perspective, theoretical consistency and the likely interaction between monetary policy and economic developments become crucial to any model that is designed to analyse monetary policy.

To provide policy makers with a useful tool to assess current economic conditions and short-term developments, Norges Bank has for the last couple of years followed a strategy of combining density forecasts from many models, see Bjørnland et al. (2008) for a description. By combining forecasts from many models, we try to hedge against uncertain instabilities.[1] The use of density forecasts is motivated by the fact that policy makers' loss function may not be quadratic or that economic developments do not follow linear trends. Thus, it no longer suffices to focus solely on first moments of possible outcomes (point forecasts).[2]

[1]The idea of combining forecasts from different models was first introduced by Bates and Granger (1969). Their main conclusion was that a combination of two point forecasts can yield lower mean square forecasts error than either of the original forecasts when optimal weights are used. Timmermann (2006) surveys combination methods and provides theoretical rationales in favor of combination - including unknown instabilities, portfolio diversification and idiosyncratic biases.

[2]Mitchell and Hall (2005) and Hall and Mitchell (2007) provide some justification for density combi-

In this paper we first describe Norges Bank's System for Averaging Models (SAM) in more detail. Since 2008, SAM has been used to provide the Bank with model-based forecasts for Mainland GDP and consumer prices adjusted for taxes and without energy (CPIATE). SAM has been developed further since the start, in particular with the introduction of a two-stage weighting scheme in August 2009, see Gerdrup et al. (2009),[3] and the introduction of real-time data vintages to estimate and evaluate models from March 2011. Furthermore, we have introduced more models with timely information from surveys and financial markets.

Second, we run an exercise where we update SAM several times during the quarter, to highlight the importance of new data releases both in terms of point and density forecasts. The exercise is done for the current and the next three quarters, which corresponds to the forecast horizon in the fan charts published on Norges Bank's web site.[4] We show that the performance of SAM improves steadily as new information arrives during the quarter.[5] Interestingly, the weights attached to the different models change during the quarter in correspondence with new data releases. In this way, our combination procedure attaches a higher weight to models with new and relevant information. The model combination is thus performing well. This results is robust to different benchmark vintages and different weighting schemes for the density forecast combination. We also show that a strategy of trying to pick the best model (selection), ex ante, is inferior to model combination.

The rest of the paper is organized as follows: In the next section we first describe

nation.

[3]See also Amisano and Geweke (2009) and Bache et al. (2011) for other references to this type of weighting scheme.

[4]Aastveit et al. (2011) run a similar exercise on U.S. real-time data for GDP, but focusing on density *nowcasting*.

[5]We highlight the importance of using non-synchronous data releases (jagged edge problem) for nowcasting in the spirit of Evans (2005), Giannone et al. (2008) and Aruoba et al. (2009). However, in SAM we differ from all these approaches in that we focus on density nowcasts in addition to point nowcasts, as well as combining nowcasts from several different models instead of relying on one specific model. We also have a longer forecast horizon.

the main output from SAM that are published in conjunction with every monetary policy meeting. In the third section we describe the modeling framework and discuss the rationale for combining densities for different model classes. The fourth section describes the recursive forecasting exercise. The results of the out-of-sample forecasting experiment as well as the results from robustness checks, are discussed in section 5. Finally, we provide a summary in section 6.

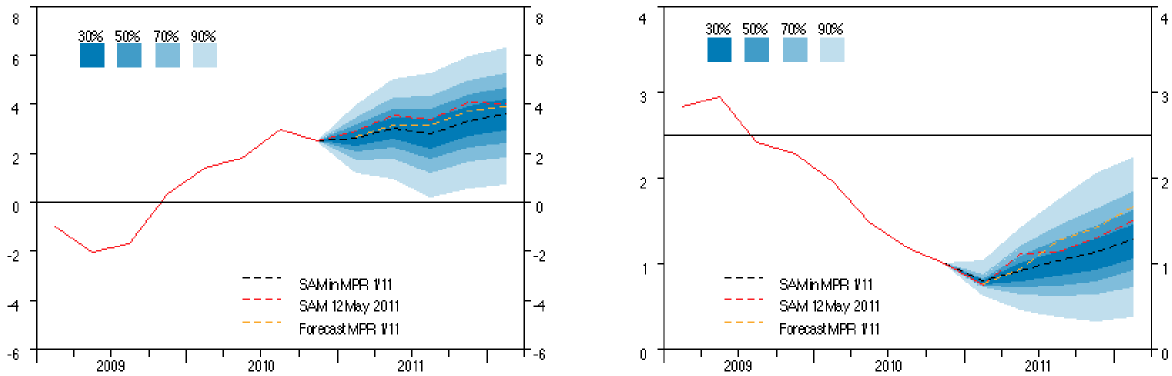
2 Communicating the main output

SAM provides model-based forecasts, but the final short-term forecasts that are published in the triannual Monetary Policy Reports (MPR) are in general subject to judgment. The final short-term forecasts are used as starting values and conditional assumptions in the Bank's core macroeconomic model, NEMO (Norwegian Economy Model).[6] SAM forecasts are updated regularly and are published on the Bank's website in conjunction with each monetary policy meeting in Norges Bank's Executive Board. Figure 1 depicts the fan charts for GDP and CPIATE from SAM published on Norges Bank's web site 12 May 2011. When MPR 1/11 was published in March, inflation and GDP were judged to increase somewhat more that the mean of the SAM densities. At the monetary policy meeting in May, SAM forecasts for growth in GDP were revised upwards. SAM forecasts for inflation were revised down from 2011Q3 onward.

The current version of SAM produces forecasts for two variables, but the aim is to produce density forecasts for most of the observable endogenous variables in NEMO.[7] In the next sections we will describe the building blocks of SAM and the performance of the system.

[6]See Brubakk et al. (2006) for a documentation of NEMO.

[7]A pilot version of SAM for private consumption already exists.



(a) GDP

(b) CPI-ATE

Figure 1. *Density forecasts for Mainland GDP and CPI-ATE from SAM and point forecasts from MPR 1/11. Four-quarter growth. Per cent*

3 Forecast methods

3.1 Component models and model classes

The forecasts produced by SAM are combinations of density forecasts for quarterly growth in Mainland GDP and four-quarter growth in CPIATE, on the basis of the flow of information that becomes available during the quarter. In practice, policymakers are often provided with forecasts from different models. For short-term forecasting, there are in particular three classes of models that are widely used: Vector autoregressive (VAR) models, leading indicator models and factor models. VAR models, first introduced by Sims (1980), are arguably the most commonly used model class for economic analysis and forecasting by policymakers. Further, there is a large amount of studies showing that leading indicators are useful for economic forecasting, see among others Banerjee et al. (2005), Banerjee and Marcellino (2006) and Marcellino (2006) for a survey on the use of leading indicators in macroeconomics. Finally, factor models have been increasingly popular at central banks as they tend to have good forecasting properties, benefitting from exploiting information from large datasets, see among others Stock and Watson

(2002), Giannone et al. (2008) and Aastveit and Trovik (2011).

There is considerable uncertainty regarding specifications, such as choosing lag lengths, data-sample, variables to include etc. for any model. Recent work by Clark and McCracken (2009) and Clark and McCracken (2010) show that for instance VARs may be prone to instabilities, and they suggest combining forecasts from a wide set of VARs. The same arguments may also apply to factor models and leading indicator models. In particular the number of factors and the choice of a stable leading indicator over a long time horizon are issues of concern.

To utilize the gains from forecast combination without being influenced by the number of models within each class, we choose to combine forecasts in two steps.[8] The density forecasts for each individual component model within a model class are combined in the first step.[9] This yields a single, combined predictive density for each model class. An advantage of this approach is that it explicitly accounts for uncertainty about model specification and instabilities within each model class. Hence, our predictive densities for each model class will be more robust to miss-specification and instabilities than following a common approach where only one model from each model class is used. In the second step, we combine the density forecasts from each model class and obtain a single combined density forecast.

Table 1 gives an overview over all models and model classes that are included in SAM. A more thorough description of the different models is left to the appendix, see section B. Below we will describe in more detail how we aggregate the density forecasts.

3.2 Combining density forecasts

In the following, we will explain how the predictive densities are combined. First we discuss the method of aggregation to use, i.e. the functional form of combining. After a

[8]See Gerdrup et al. (2009) for more details on this.

[9]Our approach is close to Aiolfi and Timmermann (2006) in the sense that we combine models in more than one stage. They find that forecasting performance can be improved by first sorting models into clusters based on their past performance, second by pooling forecasts within each cluster, and third by estimating optimal weights on these clusters (followed by shrinkage towards equal weights).

Table 1. Overview of all component models and model classes

Model class	**Models**	**Description**	**Number GDP**	**Number CPIATE**
VAR	Classical VARs	ARs and VARs using GDP (and CPIATE and/or interest rate),		
		and VARs using CPIATE (and GDP and/or interest rate)	144	156
		Lag length: $1-4$		
		Transformations: First differences, double differences, detrended		
		Estimation period: Recursive from 1992Q1 and		
		rolling (30 and 40 quarters for VAR, 20 and 40 quarters for AR)		
	Bayesian VARs	Bivariate VARs using GDP and inflation	3	3
		Transformations: First differences		
		Estimation period: Recursive from 1993Q1		
	VECM	Vector Equilibrium Model	1	1
		Estimation period: Recursive from 1982Q4		
	DSGE	Dynamic stochastic general equilibrium model	1	1
		Estimation period: Recursive from 1990Q1		
	Combination	**Linear opinion pool**		
		Log-score weights (GDP) or MSE-weights (CPIATE)	**149**	**161**
Indicator	Bivariare VARs	Bivariate VARs with GDP or inflation and different indicators		
		incl. a trivariate VAR with GDP and monetary agg. (M1,M2)	64	4
		Estimation Period: Recursive from 1992Q1 if available		
	One-eq. model	Indicator models using monthly information	3	0
		Estimation period: Recursive from 1989Q2		
	Disagg	AR-model based on forecasts for main components in CPI.		
		Forecasts for components are combined using CPI-weights	0	1
		Estimation period: Recursive from 1991Q1		
	Combination	**Linear opinion pool**		
		Log-score weights (GDP) or MSE-weights (CPIATE)	**67**	**5**
Factor	Monthly factors	Dynamic factor models	4	4
		Number of factors: $1-4$		
		Estimation period: Recursive from 1990Q1		
	Quarterly factor	Dynamic factor models using quarterly information	1	1
		Estimation period: Rolling 60 quarters		
		Number of factors: 1		
	Combination	**Linear opinion pool**		
		Log-score weights (GDP) or MSE-weights (CPIATE)	**5**	**5**
SAM	**Combination**	**Linear opinion pool**		
		Log-score weights (GDP) or MSE-weights (CPIATE)	**221**	**171**

brief explanation of our preferred evaluation criteria for predictive densities, we explain the construction of the weights attached to each model. In our two-stage approach, we choose the same method of aggregation and construction of weights for both stages. We conclude this section with some remarks on evaluating density forecasts.

3.2.1 Method of aggregation

One popular approach to solve the aggregation problem is to take a linear combination of the individual density forecasts, the so-called linear opinion pool:

$$p(y_{\tau,h}) = \sum_{i=1}^{N} w_{i,\tau,h}\, f(y_{\tau,h}|I_{i,\tau}), \quad \tau = \underline{\tau}, ..., \overline{\tau} \tag{1}$$

where N denotes the number of models to combine, $I_{i,\tau}$ is the information set used by model i at time τ to produce the density forecast $f(y_{\tau,h})$ for variable y at forecasting horizon h. $\underline{\tau}$ and $\overline{\tau}$ denote the first and last period, respectively, over which the individual models' densities are evaluated, and $w_{i,\tau,h}$ are a set of non-negative weights that sum to unity (see section 3.2.2).

Combining the N density forecasts according to equation 1 can potentially produce a combined density forecast with characteristics quite different from those of the individual models. As Hall and Mitchell (2007) notes; if all the forecast densities are normal, but with different mean and variance, then the combined density forecast using the linear opinion pool will be mixture normal. This distribution can accommodate both skewness and kurtosis and be multimodal, see Kascha and Ravazzolo (2010). If the true unknown density is non-normal, this is an appealing feature. Further, since the combined density is a linear combination of all the individual models' densities, the variance of the combined density forecast will in general, and more realistic, be higher than that of individual models' density forecast. The reason is that the variance of the combination is equal to the weighted sum of a measure of model uncertainty and dispersion (or disagreement) of the point forecasts, see Wallis (2005). If, on the other hand, the true unknown density is normal, combining the individual forecast densities using equation 1 will in general get the distribution wrong. For a discussion of alternative combination methods, see among others Bjørnland et al. (2011) and Wallis (2011).

3.2.2 Deriving the weights

Many different weighting schemes have been proposed in the literature. Equally-weighted combinations have been found to be surprisingly effective for point forecasting, see Clemen (1989) and Stock and Watson (2004). Bates and Granger (1969) propose another alternative, combining models using weights derived from their sum of squared errors (SSE). These weights will minimize a quadratic loss function based on forecast errors, provided that the estimation errors of different models are actually uncorrelated. Using inverse-SSE weights produces the same weights as those derived from the inverse of mean squared errors (MSEs) computed over some recent observed sample:

$$w_{i,\tau,h} = \frac{\frac{1}{MSE_{i,\tau,h}}}{\sum_{i=1}^{N} \frac{1}{MSE_{i,\tau,h}}}, \quad \tau = \underline{\tau}, ..., \overline{\tau} \tag{2}$$

where τ, h, N and i are defined above.

In a density combination setting, the range of possible weighting schemes is richer. It is possible to calculate MSEs based on the means of the distributions, but it is more natural to take advantage of the full distributions, see e.g. Jore et al. (2010) and Amisano and Geweke (2009). Then the question of evaluating densities arises.

A popular statistical measure is the Kullback-Leibler divergence or Kullback-Leibler information criterion (KLIC), see Mitchell and Hall (2005), Amisano and Giacomini (2007) and Kascha and Ravazzolo (2010). The KLIC is a sensible measure of accuracy since it chooses the model which on average gives higher probability to events that have actually occurred. As argued by Mitchell and Hall (2005) the KLIC provides a unified framework for evaluating, comparing and combining density forecasts, and Mitchell and Wallis (2010) show that the KLIC can be interpreted as a mean error, similar to the use of the mean error or bias in point forecast evaluation.[10] Specifically, the KLIC distance between the true density f of a random variable y_t and some candidate density $f_i(y_t)$

[10] As discussed in Hoeting et al. (1999), the log-score is a combined measure of bias and calibration.

obtained from the individual model i is defined as

$$KLIC_i = \int f_t(y_t) \ln \frac{f(y_t)}{f_i(y_t)} dy_t = E[\ln f(y_t) - \ln f_i(y_t)], \tag{3}$$

where E denotes the expectation. The KLIC difference between two densities is then defined as

$$\begin{aligned} KLIC_j - KLIC_i &= E[\ln f(y_t) - \ln f_j(y_t)] - E[\ln f(y_t) - \ln f_i(y_t)] \\ &= E[\ln f_i(y_t)] - E[\ln f_j(y_t)] \\ &= E \ln S_i - E \ln S_j, \end{aligned} \tag{4}$$

i.e. the difference between two expected logarithmic scores (log-scores). Thus, when $E \ln S_i > E \ln S_j$, then $KLIC_i < KLIC_j$ and the candidate density $f_i(y_t)$ is the preferred density. Under some regularity conditions, $E \ln S_i$ can be estimated by the average log-score

$$\ln S_i = \frac{1}{T} \sum_{t=1}^{T} \ln f_i(y_t) \tag{5}$$

It follows from equation 4 that we do not need to know the true density in order to compare two candidate densities. When comparing density forecasts, a measure of out-of-sample performance is the (out-of-sample) log-score given by

$$\ln S_{i,h} = \frac{1}{T - h - T^S + 1} \sum_{t=T^S}^{T-h} \ln f_{t+h,t,i}(y_{t+h}), \tag{6}$$

where $f_{t+h,t,i}$ denotes a prediction of the density for Y_{t+h} conditional on some information set available at time t, and T^S and T denotes respectively the starting period for the forecasts and number of observations.

Hence, the log-score is the logarithm of the probability density function evaluated at the outturn of the forecast. Following Jore et al. (2010) we define the recursive log-score weights as:

$$w_{i,\tau,h} = \frac{exp[\sum_{\underline{\tau}}^{\tau-h} ln\ f(y_{\tau,h}|I_{i,\tau})]}{\sum_{i=1}^{N} exp[\sum_{\underline{\tau}}^{\tau-h} ln\ f(y_{\tau,h}|I_{i,\tau})]} = \frac{\ln S_{i,\tau,h}}{\sum_{i=1}^{N} \ln S_{i,\tau,h}}, \quad \tau = \underline{\tau}, ..., \overline{\tau} \tag{7}$$

where τ, h, y, N, i and $g(y_{\tau,h}|I_{i,\tau})$ are defined above. Two points are worth emphasizing about this expression. The weights are derived based on out-of-sample performance, and the weights are horizon specific.

In SAM we have chosen to apply log-score weights when forecasting Mainland GDP and MSE-weights (or more precisely, inverse MSE-weights) when forecasting CPIATE.[11] The weights are recursively updated, and thus time-varying. These weighting schemes give overall a good performance, both in terms of point and density forecasts.

3.3 Evaluating density forecasts

Corradi and Swanson (2006) provide an extensive survey of the theoretical literature on density evaluation. In general, the literature can be divided in two branches. One branch is concerned with scoring rules and distance measures, where scoring rules evaluate the quality of probability forecasts by assigning a numerical score based on the forecast and the subsequent realization of the variable, as explained in section 3.2.2.

Another common approach for evaluating density forecasts provides statistics suitable for tests of the forecast density relative to the "true" unobserved density. Following Rosenblatt (1952), Dawid (1984) and Diebold et al. (1998), we evaluate the density relative to the "true" but unobserved density using the probability integral transform (pits). The pits summarize the properties of the densities, and may help us to judge whether the densities are biased in a particular direction, and whether the width of the densities has been roughly correct on average. More precisely, the pits represent the ex-ante inverse predictive cumulative distribution, evaluated at the ex-post actual observations.

A density is correctly specified if the pits are uniform, identically and, for one-step ahead forecasts, independently distributed. We evaluate the pits in this paper on the basis of a graphical presentation. The pits can be statistically tested for uniformity and independence, but with the small sample that is currently available the power of the tests are too weak to add much to the visual inspection.

[11]The choice of MSE-weights for CPIATE is discussed in section 5.2.

4 Empirical exercise

To evaluate SAM, we perform a real-time out-of-sample forecasting exercise for quarterly growth in Norwegian Mainland GDP and four-quarter growth in CPIATE. We use real-time vintage data for the Norwegian economy, see section A in the appendix for details. A key issue in this exercise is the choice of benchmark representing the "actual" measure of GDP. Stark and Croushore (2002) suggest three alternative benchmark data vintages: the most recent data vintage, the last vintage before a structural revision (called benchmark vintages) and finally data that is released a fixed period of time after the first release. We use the fifth available release of GDP as actuals. However, our results are robust to using other vintages of GDP as actuals, including the last available vintage.

In explaining this exercise, we will concentrate on forecasts for quarterly growth in GDP.[12] The *nowcast* is a two-steps ahead forecast in the beginning of a quarter. When GDP for the previous quarter is released around 50 days into the quarter, the nowcast becomes a one-step ahead forecast (or *backcast*). In the same way, the forecast for the *next* quarter starts as a three-steps ahead forecast and turns into a two-steps ahead forecast when GDP for the previous quarter is released and a one-step ahead forecast when GDP for the next quarter is released. We publish forecasts up to four-steps ahead.

The recursive forecast exercise is constructed as follows: We estimate each model on a real-time data sample and compute density forecasts. For each vintage of GDP we re-estimate all models and compute predictive densities (for all component models, model classes and the combination) for every new data release within the quarters of interest. The first vintage of GDP is from the second quarter of 2000, containing GDP for 2000Q1. Our first nowcast is then for 2000Q2, making this quarter the start of the evaluation period. The last vintage in the evaluation period was published in the fourth quarter of 2009, hence the last nowcast is for GDP-growth in 2009Q4. The fifth release of growth in this quarter was published in February 2011.

[12] The exercise for CPIATE is done in the same way.

The data we consider are either of monthly or quarterly frequency. Hence, some data will be updated every month while others are only updated once every quarter. Series such as equity prices, dividend yields, currency rates, interest rates and commodity prices are constructed as monthly averages of daily observations. Following the standard approach, data series that have similar release dates and are similar in content are grouped together in blocks.

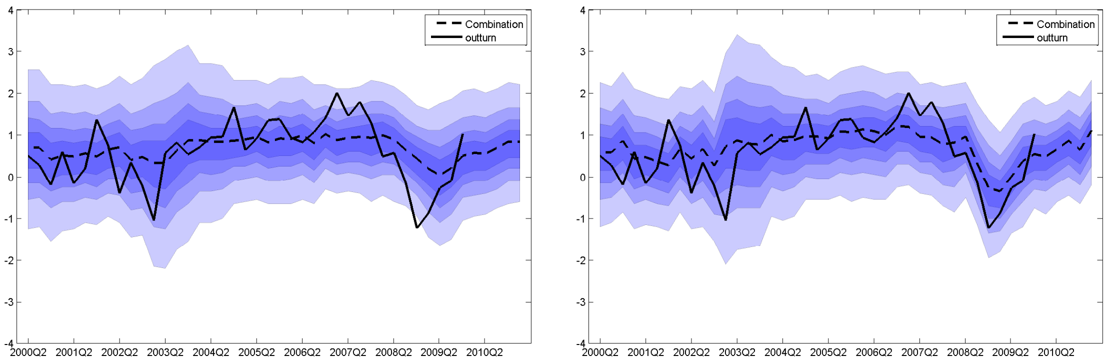
In Table C.2 in Appendix C we illustrate the data release calender (taking the nowcast as an example) that we have constructed for this exercise. The table shows, for each model class, the number of component models that update their forecast after every new data release. For example, when the Business Tendency Survey (BTS) is updated, 33 indicator models and 1 factor model are updated. It also illustrates at each point in time when the GDP and CPIATE nowcast, respectively, is a two-step ahead or a one-step ahead forecast. Note that since all the component models in the VAR class for GDP are of quarterly frequency, their forecasts only change three times per quarter. That is whenever a full quarter of CPI inflation, interest rates or GDP is available. Forecasts for the Factor model class are on the other hand updated for every monthly data release, and for some quarterly data releases.

5 Results

In this section, we analyze the performance of SAM. First, we discuss the importance of new information in terms of providing more accurate density and point forecasts, both for the current quarter and the next three quarters. Second, we inspect the different densities to see if they appear to be well-calibrated in terms of the pits. Third, we compare the performance of the three model classes that are combined in SAM. Finally, we check for robustness of our results, both with respect to selection of benchmark vintage for GDP and with respect to alternative weighting schemes.

5.1 Evaluation of results

We measure the forecasting accuracy in terms of evaluating the log-score and RMSE of the predictive densities after every new data release during a quarter. In essence, our exercise can be illustrated by figure 2 and 3 for Mainland GDP and CPIATE, respectively.

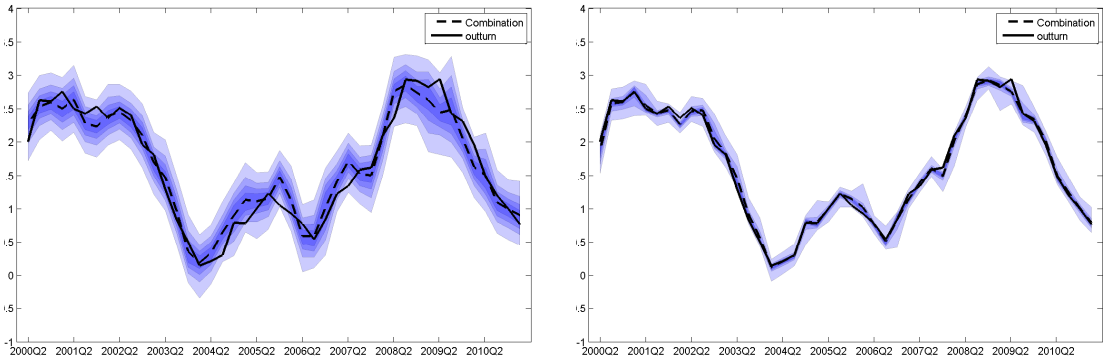


(a) Beginning of quarter ($M0$)

(b) Last day of fourth month ($M4$)

Figure 2. *Recursive real-time out-of-sample density nowcast/backcast for Mainland GDP made at two different points in time. Quarterly growth. Per cent.* $M0$ *refers to nowcasts made at the first day of the representing quarter, while* $M4$ *refers to the backcast made at the last day of the fourth month, i.e. around three weeks before the publication of GDP. The solid line shows the fifth release of GDP. The shaded blue areas are, from darkest to lightest, 30%, 50%, 70% and 90% probability bands.*

The figures shows recursive real-time out-of-sample density nowcasts for quarterly growth in GDP and four-quarter growth in CPIATE for the period 2000Q2-2011Q1 and outturns (fifth release for GDP). The nowcasts are made at two different points in time during the quarter. Recursive nowcasts for GDP growth made on the first day of the quarter are shown in the left panel in figure 2, while recursive nowcasts made on the last day of the fourth month after the beginning of exercise, i.e. backcast made around three weeks prior to the release of Mainland GDP, are shown in the right panel. For CPIATE, recursive forecasts are made on the first day of the quarter and on the last day of the third month after the beginning of exercise, 9 days before the release of the full quarter CPIATE.



(a) Beginning of quarter ($M0$)

(b) Last day of third month ($M3$)

Figure 3. *Recursive real-time out-of-sample density nowcast/backcast for CPIATE made at two different points in time. Four-quarter growth. Per cent.* $M0$ *refers to nowcasts made at the first day of the representing quarter, while* $M3$ *refers to backcast made at the first day of the fourth month, i.e. around 9 days before the publication of CPIATE for the whole quarter. The solid line shows the actuals. The shaded blue areas are, from darkest to lightest, 30%, 50%, 70% and 90% probability bands.*

The exercise enables us to study how the predictive densities change as more data are available throughout the quarter, i.e. for each block of information as shown in table C.2 in Appendix C. Furthermore, we extend the analysis to cover up to three quarters ahead. We want to evaluate both the accuracy of our density forecasts and how well they are calibrated. From the two panels in figures 2 and 3 we can see how the predictive densities are better centered around the outturn when more information is available. This may indicate that more information improves the density forecasts in terms of a higher log-score and lower RMSE. However, the density forecasts for Mainland GDP appear to be too wide, since there are very few outcomes in the 70-90 per cent band. This will be investigated further when we look at the pits.

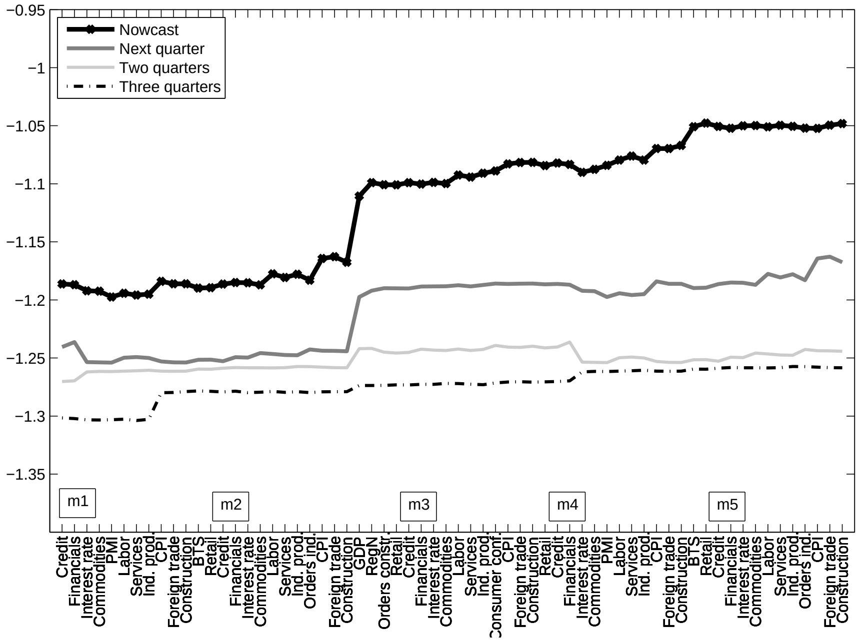


Figure 4. *Average log-scores for forecasts of quarterly growth in Mainland GDP after different block releases and for different horizons. Evaluated against 5th release of outcomes.*

Figures 4 and 5 summarize the results for Mainland GDP and CPIATE, respectively, for all horizons (four including the nowcast). The figures shows three interesting results. First, uncertainty increases as we forecast longer horizons, since the log-scores of the predictive densities are lower as we increase the horizon. Second, the forecasting accuracy improves when new information becomes available. The log-score of the predictive densities typically increases as new information arrives during the quarter. Publication of GDP for the previous quarter have the biggest impact on the forecast accuracy for the current and next quarter. When forecasting CPIATE, it is more or less only monthly realizations of CPIATE itself, and to some extent GDP, that improves forecast accuracy at all horizons. Third, for GDP new information has a much larger impact on the accuracy of the nowcast and the next quarter than on the accuracy of the forecasts for the last two quarters. The improvement in accuracy when new information arrives is small when forecasting two quarters ahead and practically non-existent when forecasting three quarters ahead. Not even the publication of GDP for the previous quarter has much impact on forecast accuracy when forecasting two or three quarters ahead.

We get the same results when evaluating the point forecasts (i.e the midpoint of the

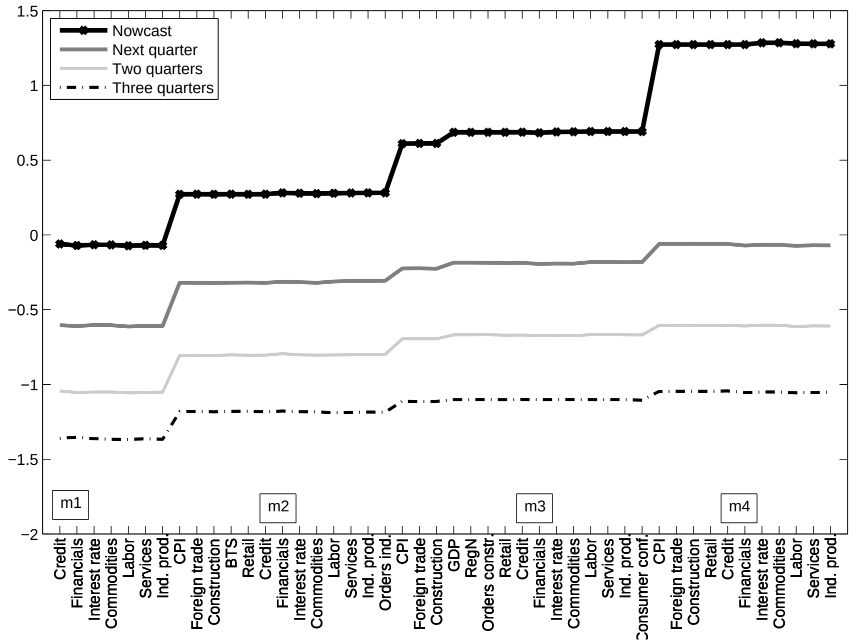


Figure 5. *Average log-score for forecasts of four-quarter growth in CPIATE after different block releases and for different horizons.*

distribution), see figures C.1 and C.2 in appendix C.

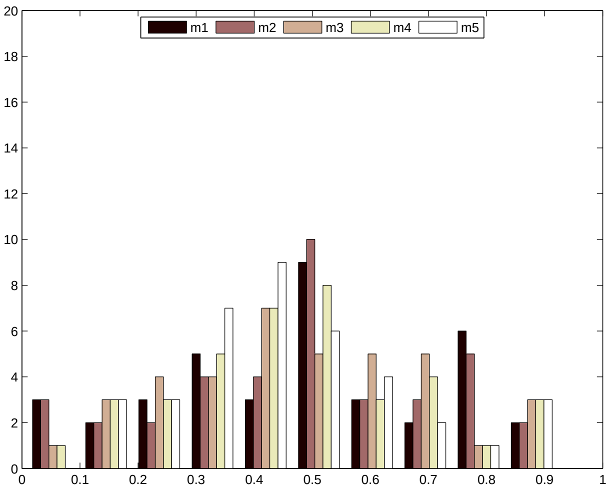
5.2 Evaluating pits

We evaluate the predictive densities relative to the "true", but unobserved, density, using the pits of the realization of the variable with respect to the forecast densities at particular points in time during the quarter. See figure 6 and 7 for Mainland GDP and CPIATE, respectively. The pits show, loosely speaking, the number of times outturns lie in the different parts of the density. A density may be correctly calibrated if the pits are uniformly distributed.

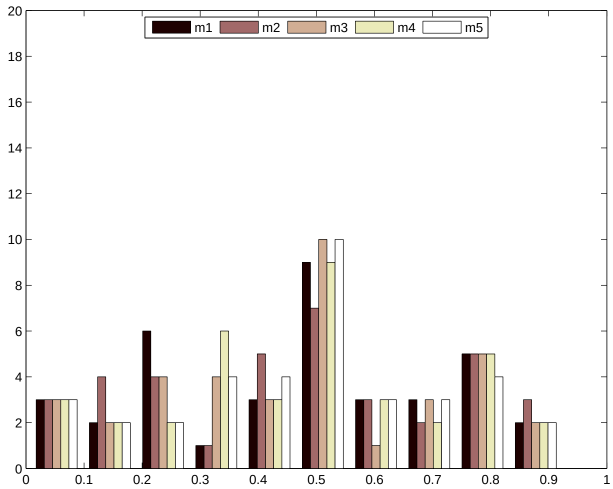
The density forecasts for Mainland GDP appear to be somewhat wide since we have a clustering of outturns in the central part of the densities. Furthermore, no outturns fall in the upper 10 per cent part of the densities when forecasting the current quarter, next quarter or two quarters ahead. The reason for this result is that models are estimated on a series for GDP that are quite volatile (the full vintage of GDP that are available at each point in time), but evaluated against a more smooth series for GDP (the fifth release), see also figure C.3 in Appendix C. The pits are even more concentrated when we evaluate against the first release of Mainland GDP, as shown in figure C.7 in Appendix

C, since this series is even more smooth than the fifth release. The pits for Mainland GDP appear to be much better calibrated when evaluated against the last vintage of Mainland GDP, see figure C.8 in Appendix C, but at the cost of less precise forecasts in terms of log-score and RMSE.

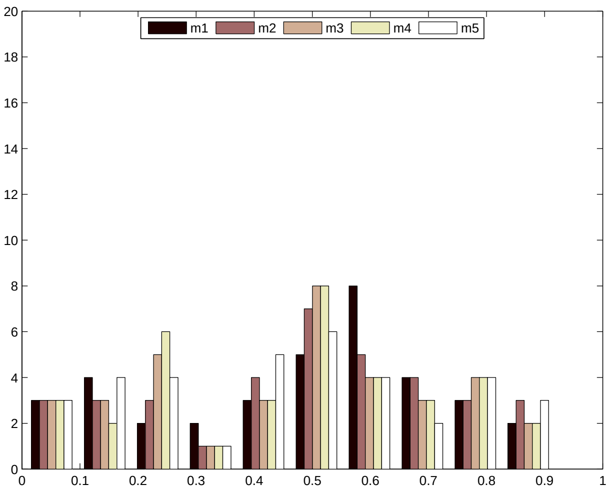
The pits for the CPIATE shown in figure 7 appear to be reasonably well-calibrated, and the choice of MSE-weights in SAM was indeed based on this result. Arguably, it appears more consistent to use log-score weights also for CPIATE, since this weighting scheme attach more weight to the best performing densities. However, when evaluating density forecast out-of-sample, there is some evidence that employing log-score weights for CPIATE leads to too narrow densities. There is a more pronounced clustering of pits at the lower end of the density forecasts using log-score weighting, see figure C.4 in Appendix C. However, the differences in pits between the preferred one in SAM and those derived from log-score weights are not large. To summarize this point, MSE-weights for CPIATE entail somewhat better calibrated density forecasts while not being inferior to alternative weighting schemes in terms of RMSE and log-score.



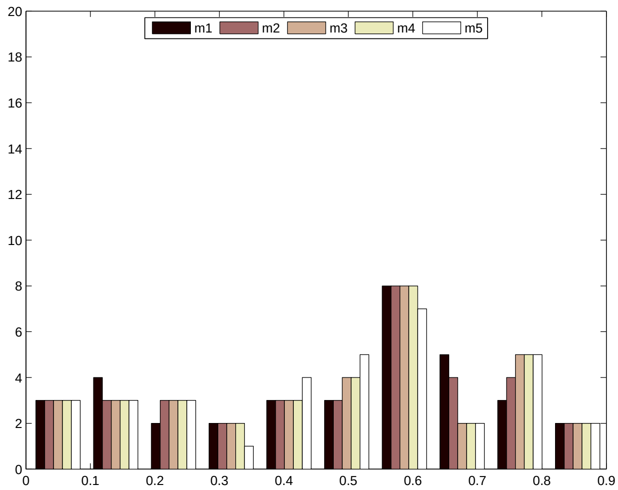
(a) Nowcast



(b) Next quarter

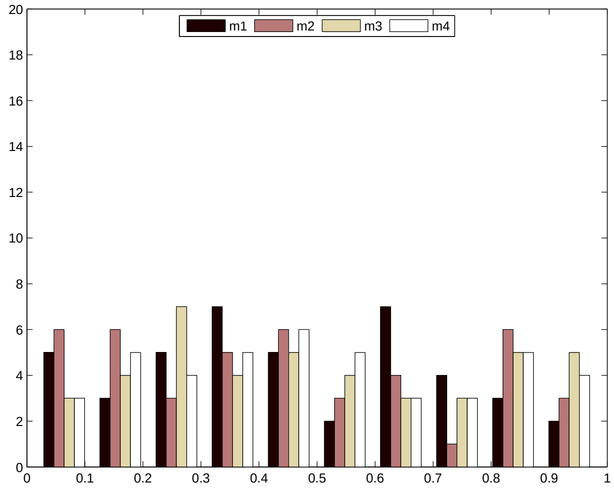


(c) Two quarters

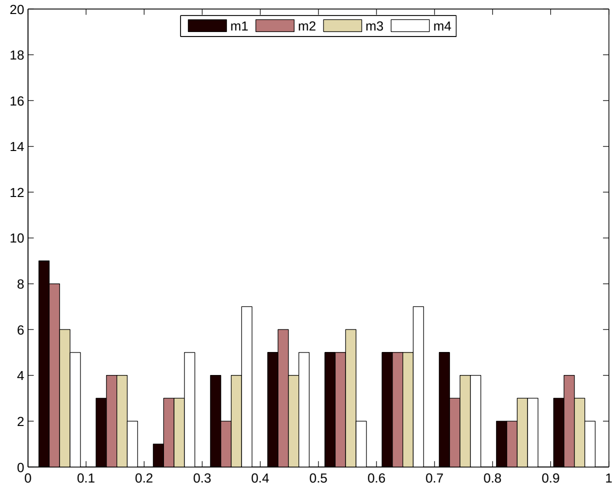


(d) Three quarters

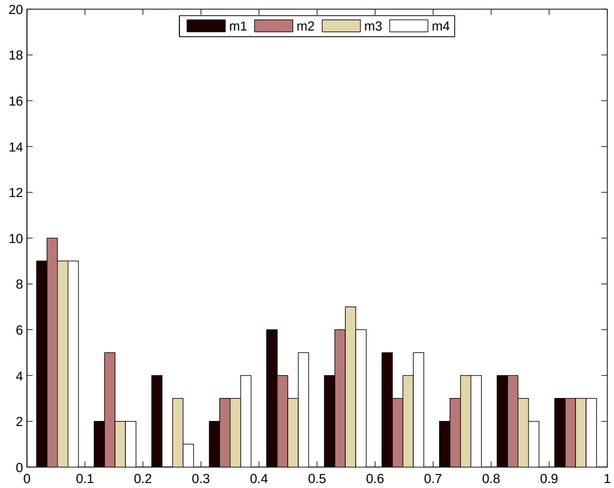
Figure 6. *Probability integral transforms (pits) at different points in time. Forecasts of quarterly growth in Mainland GDP. The pits of a forecasting model should have a standard uniform distribution if the model is correctly specified.*



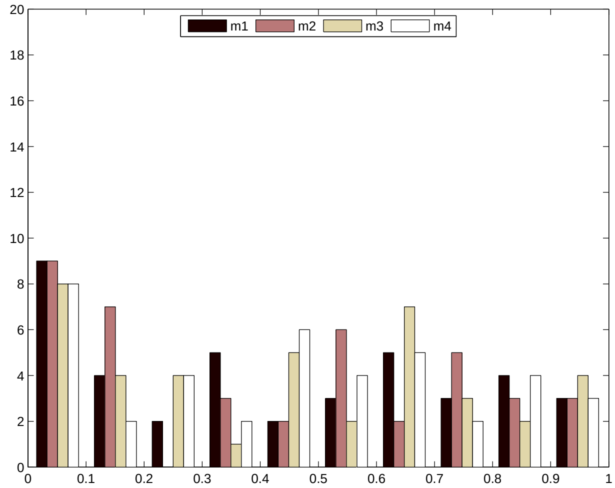
(a) Nowcast



(b) Next quarter



(c) Two quarters



(d) Three quarters

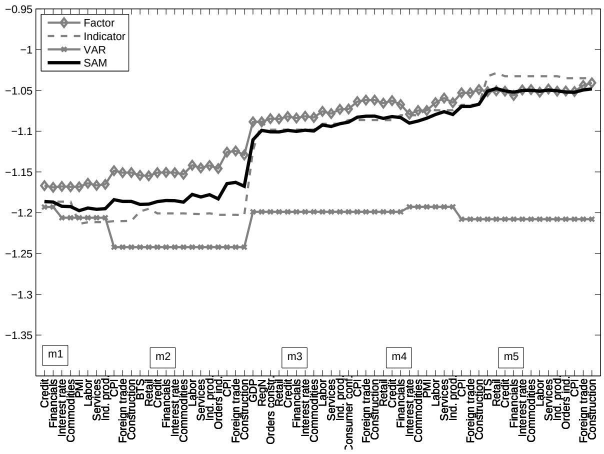
Figure 7. *Probability integral transforms (pits) at different points in time. Forecasts of four-quarter growth in CPIATE. The pits of a forecasting model should have a standard uniform distribution if the model is correctly specified.*

5.3 Performance of model classes and changes of weights in SAM

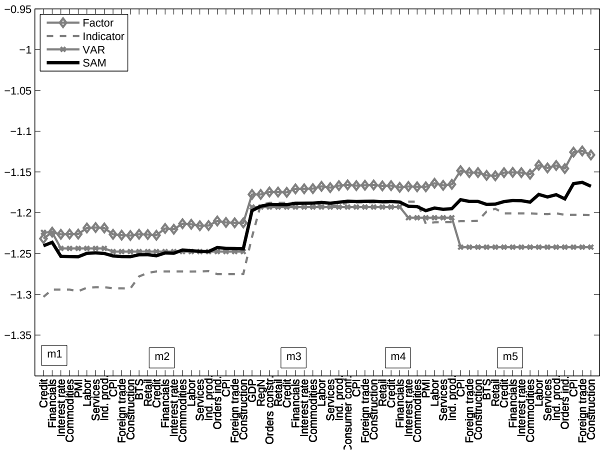
Figures 8 and 9 show results for Mainland GDP and CPIATE, respectively, in terms of average log-scores. Four results can be highlighted from the figures for Mainland GDP. First, when forecasting the current and next quarter, the forecasting performance of the Factor class, Indicator class and SAM improves when new information becomes available. Second, the Factor class outperforms the other model classes most of the time when forecasting current and next quarter's GDP. The Indicator model class has, however, the best performance for the current quarter in terms of log-score after the release of the Business Tendency Survey. Third, the VAR model class' relative performance improves greatly at longer horizons, as the informational advantage of the Factor model class and the Indicator model class become less important. This result may reflect the large number of different specifications of (V)AR models in this group. Fourth, SAM performs overall very well. This is due to weights shifting to the model class with the best performance at different points in time through all forecast horizons.

When forecasting CPIATE, however, results are somewhat different. As mentioned in section 5.1, it is mostly only monthly realizations of CPIATE itself, and to some extent GDP, that improves forecast performance at all horizons. Having a broad information set seems therefore to add little extra value to performance. The Indicator model class has the best performance most of the time and for all horizons. The most important model in that class is a bivariate VAR using monthly data of CPIATE and registered unemployment as explanatory variables, but lagged values of previous month's CPIATE are also important in this model. SAM also performs, overall, very well for CPIATE.

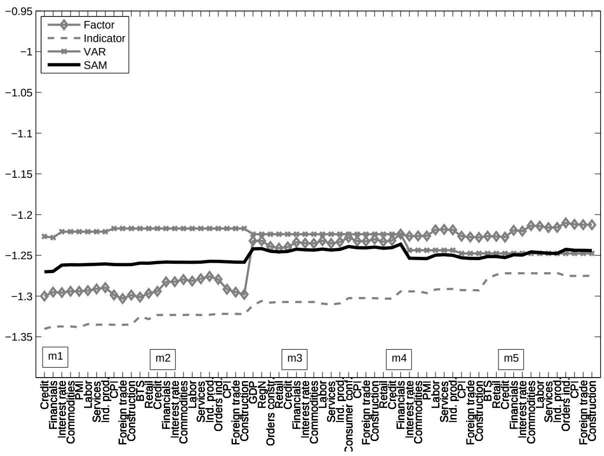
In Appendix C, figures C.5 and C.6 show results for evaluating point forecasts for Mainland GDP and CPIATE in terms of RMSE. The same conclusions apply.



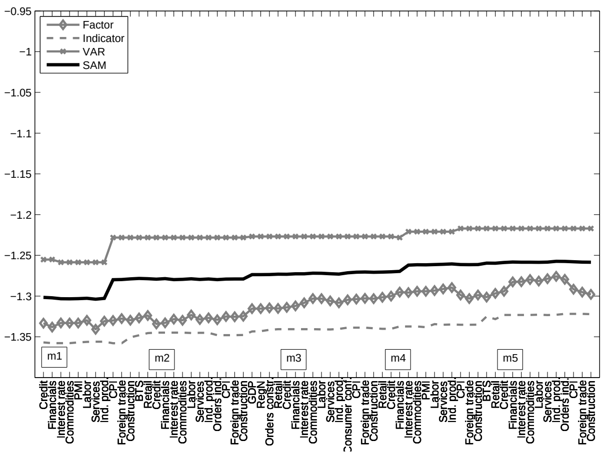
(a) Nowcast



(b) Next quarter

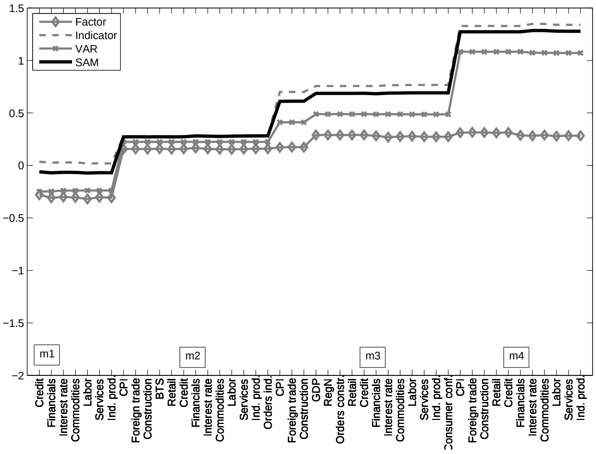


(c) Two quarters

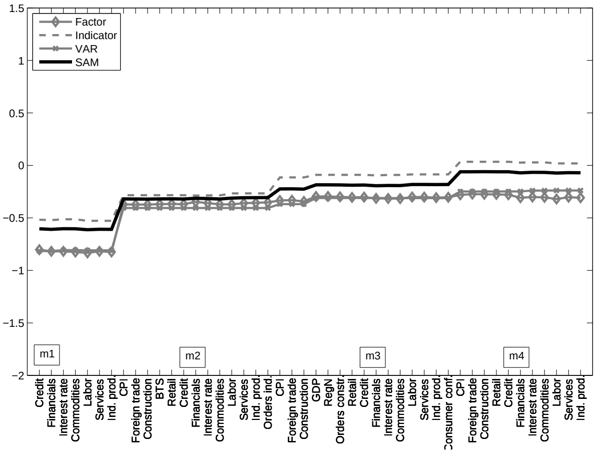


(d) Three quarters

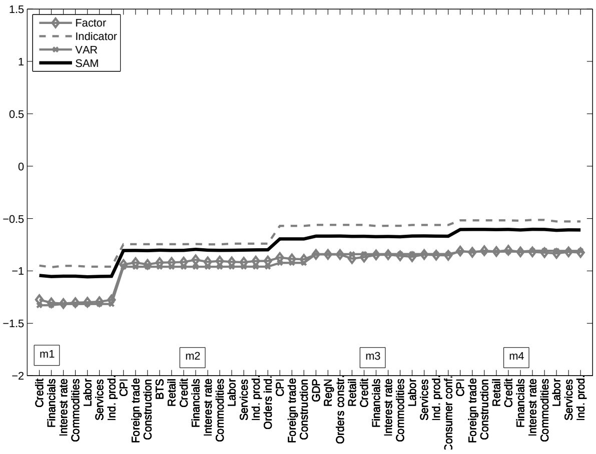
Figure 8. *Average log scores for forecasts made after different block releases for model classes and SAM for different horizons. Forecasts of quarterly growth in Mainland GDP. Evaluation against 5th releases of Mainland GDP*



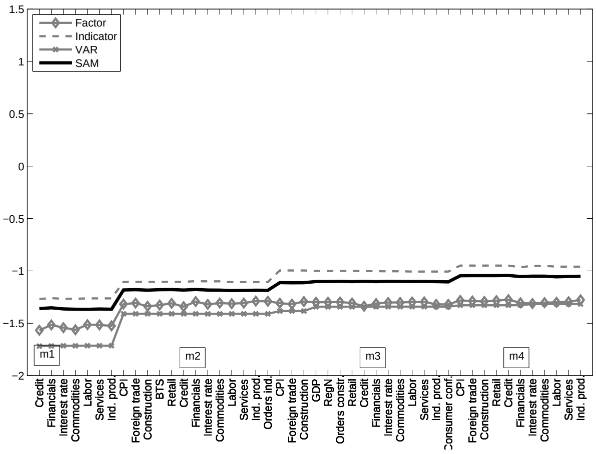
(a) Nowcast



(b) Next quarter



(c) Two quarters



(d) Three quarters

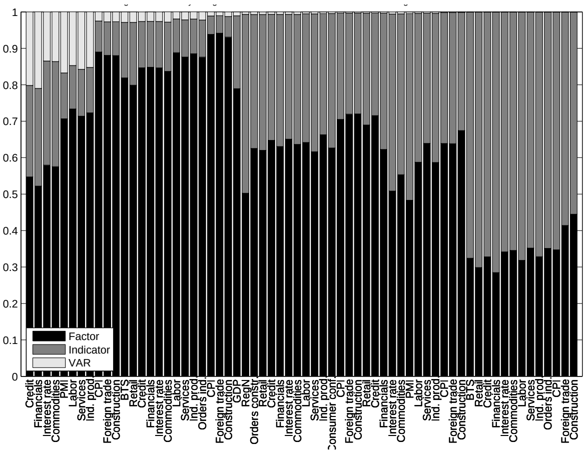
Figure 9. *Average log scores for forecasts made after different block releases for model classes and SAM for different horizons. Forecasts of four-quarter growth in CPIATE*

In figure 10 and 11 we depict the weights attached to each model class in SAM after every data block release for Mainland GDP and CPIATE, respectively.[13] The weights mirror the performance in terms of log-score (for Mainland GDP), as shown in figure 8, and RMSE (for CPIATE), as shown in figure C.6 in Appendix C. Interestingly, the figures show that there are large changes in the weights throughout the quarter. This is particularly true for GDP since log-score weights can discriminates markedly between models with different log-scores.

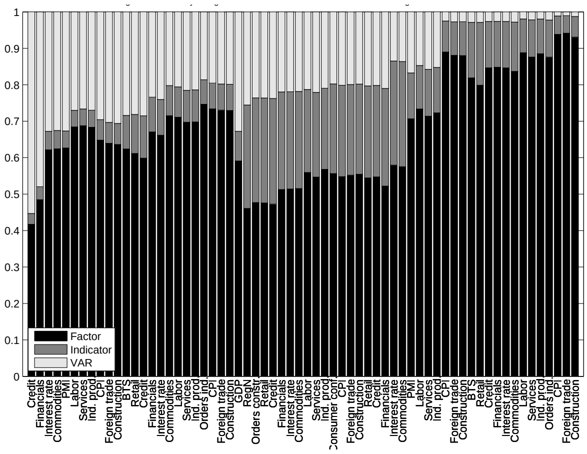
When forecasting current quarter's Mainland GDP, the VAR model class has some weight early in the quarter, but this weight is reduced to close to zero as the two other model classes can take advantage of more information. The weight on the Indicator model class is also reduced early in the quarter because of the improved performance of the Factor model class. However, after the release of GDP (for the previous quarter) and subsequently RegN (indicators from Norges Bank's regional network), the Indicator model class receives a much higher weight. Later, the release of Financials, Interest rate, PMI and in particular, the BTS (Business Tendency Survey), lead to a markedly higher weight on the Indicator model class. For the next couple of quarters, the weight on the VAR model class increases considerably. Three quarters ahead, the VAR class has a weight of more than 90 per cent most of the time. However, the weights may change depending on the choice of benchmark vintage. See section 5.4 for more details on robustness.

For CPIATE, the weight on the Indicator model class is particularly high, and increasing with the release of monthly CPI from the second month in the quarter. This reflects the high performance of this model class in terms of RMSE.

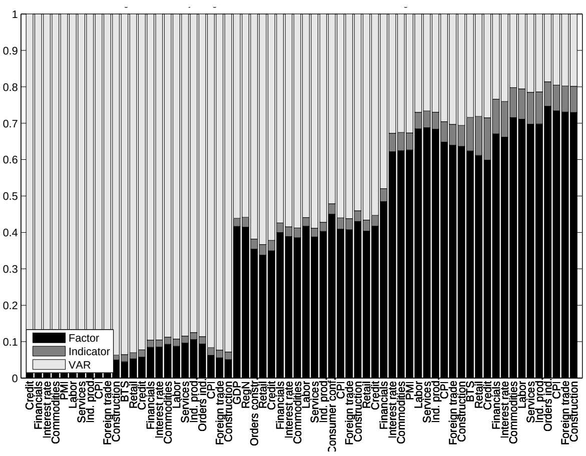
[13]The weights are calculated at the end of the evaluation period, and would typically be different in an earlier period.



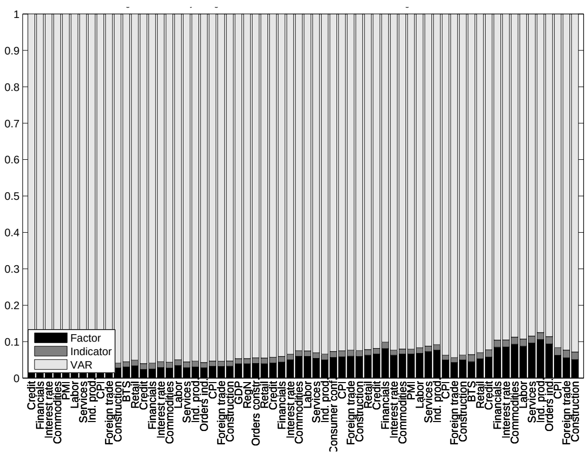
(a) Nowcast



(b) Next quarter

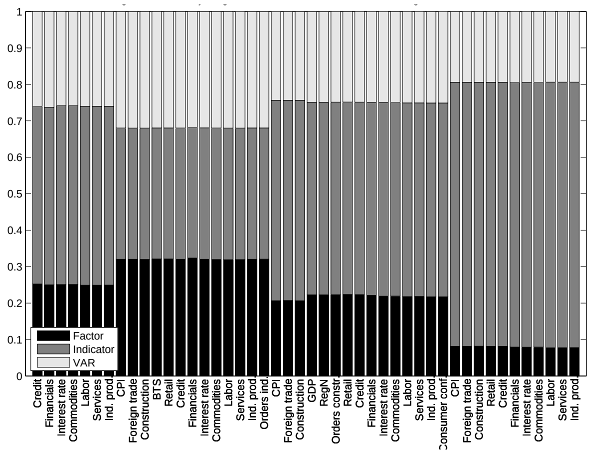


(c) Two quarters

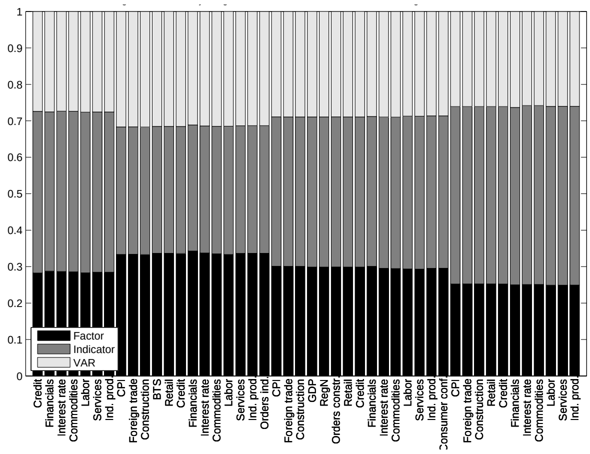


(d) Three quarters

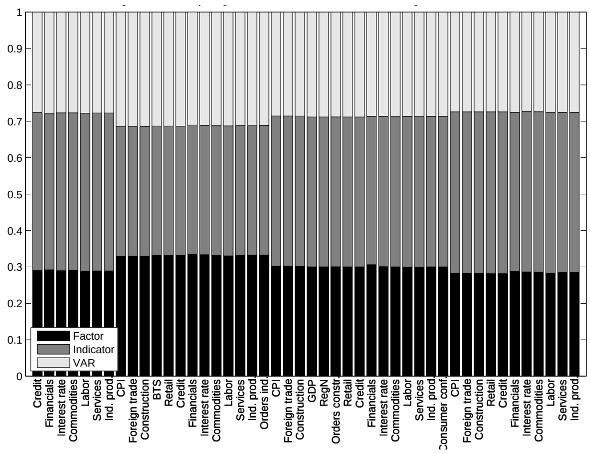
Figure 10. *Weights attached to the different model classes after different block releases. Forecasts for quarterly growth in Mainland GDP. Evaluation against 5th release of Mainland GDP*



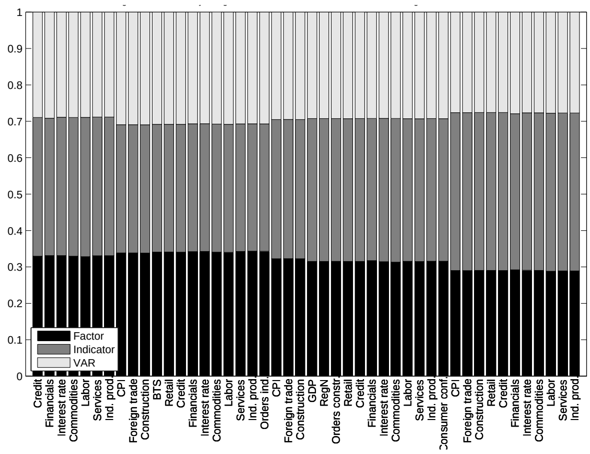
(a) Nowcast



(b) Next quarter



(c) Two quarters



(d) Three quarters

Figure 11. *Weights attached to the different model classes after different block releases. Forecasts for four-quarter growth in CPIATE*

5.4 Robustness

We check for robustness of our results in two ways. First, with respect to selection of benchmark vintage for GDP. Second, we check for robustness with respect to alternative weighting schemes for both GDP and CPIATE. For simplicity, we only focus on robustness results for the nowcasts.

5.4.1 Alternative benchmark vintages

The choice of benchmark vintage is a key issue in any application using real-time vintage data.[14] In SAM, we have chosen to use the fifth available vintage of Mainland GDP as benchmark, ie growth rates have been revised four times. Figure C.7 and C.8 in Appendix C shows out-of-sample results with the first release of GDP and the last available vintage of GDP as benchmark, respectively. The two figures show clearly that there are differences in terms of nowcasting performance from the different model classes, depending on the benchmark vintage. Hence, also the weights attached to the different model classes differ. However, the result that the combined nowcast is always performing well seems to be remarkably robust. This may imply that there are additional gains from combining forecasts in a real-time environment where the forecast target (benchmark) is not obvious.

5.4.2 Alternative weighting schemes for the combination

Several papers have found that simple combination forecasts, such as equal weights, outperform more sophisticated adaptive forecast combination methods. This is often referred to as the forecast combination puzzle. While Jore et al. (2010) and Gerdrup et al. (2009) seem to find some evidence of gains from adaptive weights based on out-of-sample performance for density combination, this is still a question of debate. We check for robustness with respect to the following different weighting schemes:

[14] See Croushore (2006) for a survey on forecasting with real-time macroeconomic data

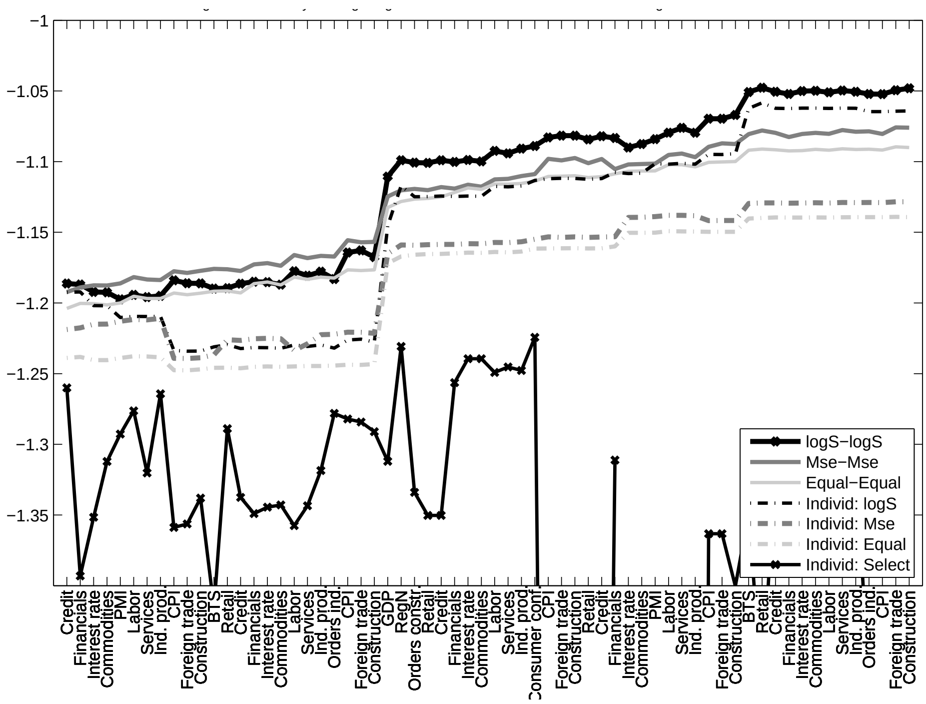


Figure 12. *Comparing different weighting schemes. Average log-scores for forecasts after different block releases. Forecasts of quarterly growth in Mainland GDP evaluated against 5th release of outcomes.*

1. Combination of all component models applying equal weights (Individ:Equal)

2. Combination of all component models applying log-score weights (Individ:logS)

3. Combination of all component models applying MSE-weights (Individ:Mse)

4. Two-stage combination with equal weights in both stages (Equal-Equal)

5. Two-stage combination with log-score weights in both stages (LogS-LogS)

6. Two-stage combination with MSE-weights in both stages (Mse-Mse)

7. Selection strategy where we try to pick the, ex-ante, "best" model at each point in time (Individ:Select) based on log-scores

Figure 12 compares the average log-scores for different weighting schemes for forecasts of GDP growth, while figure 13 compares the scores for inflation forecasts. There are several interesting results. First, all combination methods yield a steady increase in log-scores as more information becomes available. This is not the case for the selection

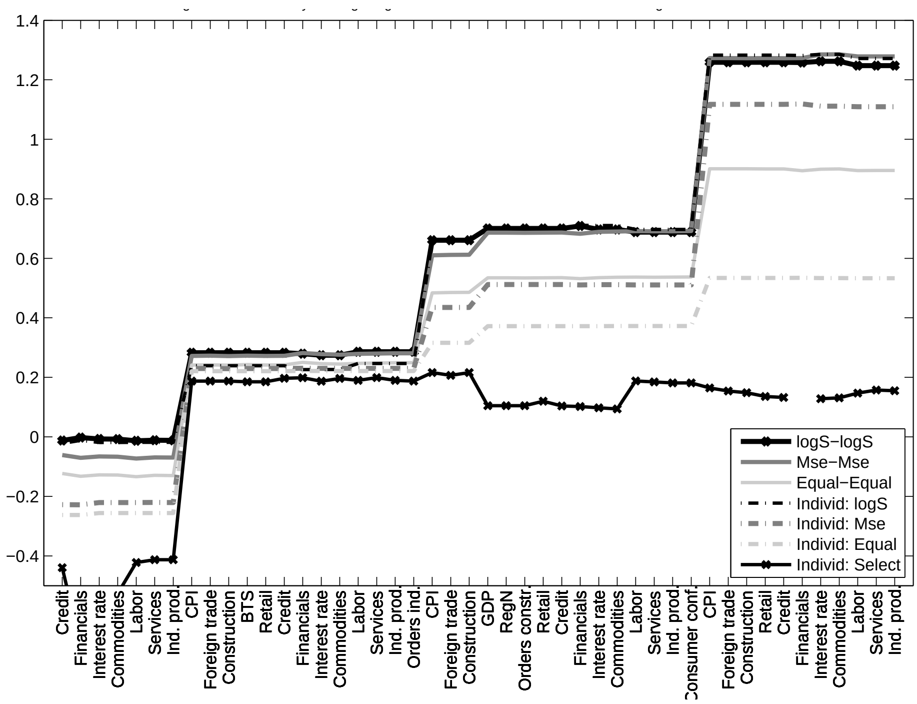


Figure 13. *Comparing different weighting schemes. Average log-scores for forecasts after different block releases. Forecasts of fourth-quarter growth in CPIATE.*

strategy, which has inferior performance in terms of low and volatile log-scores when new information arrives. Second, the difference between Individ:Equal and Equal-Equal can be seen as the "pure" gain from using a two-stage approach where models are first grouped into model classes and then combined in a second step. It is evident from the figure that for both variables, the two-step procedure Equal-Equal is always performing better than the procedure where all individual models are combined in one step using Equal weights (Individ:Equal). Third, there are small differences between Individ:LogS and LogS-LogS when forecasting CPIATE. When forecasting GDP, however, the performance of LogS-LogS and the other two-stage combination schemes are better than the one-step weighting schemes. Finally, no weighting scheme is superior throughout the quarter, but our two-stage combination approach combined with some adaptive weights, based on log-scores for GDP and MSE for CPIATE, seem to be a robust strategy for combining models.

The results are quite similar when comparing RMSEs, see figure C.9 in Appendix C.

6 Summary

In this paper we describe Norges Bank's 'System for averaging models' (SAM), which produces model-based density forecasts for quarterly growth in Mainland GDP and four-quarter growth in CPIATE. Forecasts for Mainland GDP are evaluated against the fifth release of GDP, since this release incorporates important revisions without adding too much volatility as in more final historical vintages.

We combine the forecasts from three main types of models typically used at central banks: Vector autoregressive models, leading indicator models and factor models. To utilize the gains from forecast combination without being influenced by the number of models within each class, we choose to combine forecasts in two steps. The density forecasts for each individual component model within a model class are combined in the first step. This yields a single, combined predictive density for each model class. An advantage of this approach, is that it explicitly accounts for uncertainty about model specification and instabilities within each model class. In the second step, we combine the density forecasts from each model class and obtain a single combined density forecast.

We update SAM several times during the quarter to highlight the importance of new data releases for forecasts for the current and next three quarters. We show that the performance of SAM improves steadily as new information arrives during the quarter. However, the value of new information is reduced as we extend the forecast horizon, in particular when forecasting Mainland GDP. Furthermore, we show that SAM is robust with regard to alternative vintages of data to evaluate against and that other weighting schemes do not yield better performance. Finally, we show that a strategy of trying to pick the best model, ex ante, is inferior to model combination.

References

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud (2011). Nowcasting GDP in Real-Time: A Density Combination Approach. Working Paper forthcoming, Norges Bank.

Aastveit, K. A. and T. G. Trovik (2011). Nowcasting norwegian GDP: The role of asset prices in a small open economy. *Empirical Economics forthcoming.*

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*(1-2), 31–53.

Akram, Q. F. (2008). The Econometric Model of Mainland Norway (EMod) on seasonally adjusted data. Mimeo, Norges Bank.

Amisano, G. and J. Geweke (2009). Optimal prediction pools. Working Paper Series 1017, European Central Bank.

Amisano, G. and R. Giacomini (2007). Comparing density foreasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics 25*(2), 177–190.

Angelini, E., G. Camba-Mendez, D. Giannone, L. Reichlin, and G. Rünstler (2011, February). Short-term forecasts of euro area GDP growth. *Econometrics Journal 14*(1), C25–C44.

Aruoba, S. B., F. X. Diebold, and C. Scotti (2009). Real-time measurement of business conditions. *Journal of Business & Economic Statistics 27*(4), 417–427.

Bache, I. W., A. S. Jore, J. Mitchell, and S. P. Vahey (2011). Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control forthcoming.*

Baffigi, A., R. Golinelli, and G. Parigi (2004). Bridge models to forecast the euro area GDP. *International Journal of Forecasting 20*(3), 447–460.

Banerjee, A. and M. Marcellino (2006). Are there any reliable leading indicators for US inflation and GDP growth? *International Journal of Forecasting 22*(1), 137–151.

Banerjee, A., M. Marcellino, and I. Masten (2005). Leading Indicators for Euro-area Inflation and GDP Growth. *Oxford Bulletin of Economics and Statistics 67*(s1), 785–813.

Bates, J. and C. Granger (1969). The combination of forecasts. *Operations Research Quarterly 20*(4), 451–468.

Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2008). Improving and evaluating short term forecasts at the Norges Bank. Staff Memo 2008/04, Norges Bank.

Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2011). Weights and pools for a Norwegian density combination. *The North American Journal of Economics and Finance 22*(1), 61–76.

Brubakk, L., T. A. Husebø, J. Maih, K. Olsen, and M. Øsntor (2006). Finding NEMO: Documentation of the Norwegian economy model. Staff Memo 06/2006, Norges Bank.

Clark, T. E. and M. W. McCracken (2009). Improving Forecast Accuracy By Combining Recursive And Rolling Forecasts. *International Economic Review 50*(2), 363–395.

Clark, T. E. and M. W. McCracken (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(1), 5–29.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting 5*.

Corradi, V. and N. R. Swanson (2006). *Predictive Density Evaluation*. Elsevier.

Croushore, D. (2006). *Forecasting with Real-Time Macroeconomic Data*. Elsevier.

Dawid, A. P. (1984). Statistical theory: the prequential approach. *Journal of the Royal Statistical Society A 147*(2), 278–290.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review 39*(4), 863–83.

Evans, M. D. (2005). Where Are We Now? Real-Time Estimates of the Macro Economy. *International Journal of Central Banking 1*(2), 127–175.

Gerdrup, K. and J. Nicolaisen (2011). On the purpose of models - the Norges Bank experience. Staff Memo 06/2011, Norges Bank.

Gerdrup, K. R., A. S. Jore, C. Smith, and L. A. Thorsrud (2009). Evaluating ensemble density combination - forecasting GDP and inflation. Working Paper 2009/19, Norges Bank.

Giannone, D., L. Reichlin, and D. Small (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics 55*(4), 665–676.

Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting 23*(1), 1–13.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science 14*(4), 382–417.

Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics 25*(4), 621–634.

Kascha, C. and F. Ravazzolo (2010). Combining inflation density forecasts. *Journal of Forecasting 29*(1-2), 231–250.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis.* Springer.

Marcellino, M. (2006). Leading indicators. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 879–960. Amsterdam: Elsevier.

Mitchell, J. and S. G. Hall (2005). Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation. *Oxford Bulletin of Economics and Statistics 67*(s1), 995–1033.

Mitchell, J. and K. Wallis (2010). Evaluating density forecasts: Forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics forthcoming.*

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*(3), 470–472.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica 48*(1), 1–48.

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

Stock, J. H. and M. W. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

Stock, J. H. and M. W. Watson (2004). Combining forecasts of output growth in seven-country data set. *Journal of Forecasting 23*, 405–430.

Timmermann, A. (2006). Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 136–96. Amsterdam: Elsevier.

Wallis, K. F. (2005). Combining density and interval forecasts: A modest proposal. *Oxford Bulletin of Economics and Statistics 67*(s1), 983–994.

Wallis, K. F. (2011). Combining forecasts - forty years later. *Applied Financial Economics 21*(1-2), 33–41.

A Data

A.1 Description of the real-time dataset

We use a large data set, containing quarterly and monthly data. Most of the data are published by Statistics Norway (SN). Other sources of data are Ecowin and Datastream (financials and commodities), Norges Bank (interest and exchange rates) and the Norwegian Labour and Welfare Organisation. Several organizations supply survey data: Statistics Norway, TNS Gallup, Fokus Bank and Norges Bank.

We aim to use real-time data in the recursive forecasting exercise. Some data are heavily revised, like Quarterly National Accounts data (QNA). When new statistics are published every quarter, Statistics Norway also revise earlier published data. New information becomes available, the base year is changed every year and major revisions are undertaken occasionally. Seasonally adjusted national accounts data represent an additional challenge: Main aggregates are a result of aggregation of many components with nonhomogenous properties.[15] This entails consistency between the subcomponents and the aggregates in the preliminary data, but it also means that new vintages of historical data are not quite comparable to earlier vintages.

Figure C.3 in Appendix C illustrates revisions in quarterly growth rates for seasonally adjusted GDP. The solid line is the quarterly growth in GDP from 2000Q2 to 2011Q1, released in May 2011. The dotted red line depicts the growth rates as they were first published, i.e. the first releases. The difference between the two lines represents revisions from the first release of the data to the final vintage. Some of the revisions are caused by the method of aggregating seasonally adjusted subcomponents, and the growth rates in the period where the QNA numbers are final (before 2007) are much more volatile than the first release growth rates. The dashed line represents the fifth release of the growth rates, published one year after the first release. When we evaluate the forecasts in the recursive forecasting exercise to calculate the model weights, we use the fifth release as our "target". Then we avoid trying to forecast the volatile final historical growth rates,

[15] See Statistics Norway's QNA website.

while we have ensured that the data are revised quite substantially, incorporating news, compared to the first release.

Important monthly indicators (retail sales, industrial production and building starts) are also revised - sometimes heavily. For other types of economic data, for instance survey data and consumer prices, the unadjusted data are not revised, but changing seasonal factors entails revision for these data as well. Finally, financial, interest and exchange rate data are examples of data that are not revised, and seasonal adjustments are not used.

Since May 2009 Norges Bank has saved most of the available economic time series in real-time data bases. Earlier published main aggregates of National Accounts data has been collected from SN's website, and for these aggregates we have real time vintages of data since June 2000.[16] For the Business Tendency Survey (BTS) we have truncated the unadjusted series recursively backwards, and seasonally adjusted and smoothed these truncated series with the same methods as SN uses. For most of the other series, we have real-time data from 2009Q2. Thus we use the 2009Q2 vintage and truncate these series recursively backwards.

The full forecasting period runs from 2000Q2. Forecasts for all models for Mainland GDP through the forecasting period are evaluated against the 5th release of the respective outcomes, and the model weights are calculated recursively. Inflation forecasts are evaluated against the final vintage, since revisions of seasonal factors are minor, in particular since we evaluate against four-quarter growth in CPIATE.

Summing up, our data set contains a large number of data of mixed frequency. For some of the data (including main aggregates of National account and parts of the Business Tendency Survey) we have real-time data for the whole period. For almost all other data that are subject to revisions, we have real-time vintages from 2009Q2. An overview of all the data series and the availability of real time vintages is given in the next section.

[16]For some of the vintages only 8 or 12 observations were published. For these vintages, we extrapolated data backwards, based on growth in neighboring vintages.

A.2 Overview of all dataseries included in SAM

Block Names	Freq	Description	Start Vintage
Financials	m	House prices	Last vintage
Financials	m	NOK: Market - Total Return Index	Last vintage
Financials	m	NOK: Oil & Gast - Total Return Index	Last vintage
Financials	m	NOK: Basic Materials - Total Return Index	Last vintage
Financials	m	NOK: Industrials - Total Return Index	Last vintage
Financials	m	NOK: Consumer Services - Total Return Index	Last vintage
Financials	m	NOK: Utilities - Total Return Index	Last vintage
Financials	m	NOK: Financials - Total Return Index	Last vintage
Financials	m	NOK: Technology - Total Return Index	Last vintage
Financials	m	NOK: Liquidity measure. Relative spread (bid/ask). All companies	Last vintage
Financials	m	NOK: Amihuds illiquidity ratio - all companies	Last vintage
Interest Rates	m	NOK: Interest rate. Interbank 3 month - offered rate	Last vintage
Interest Rates	m	NOK: Interest rate. Interbank 6 month - offered rate	Last vintage
Interest Rates	m	NOK: Interest rate. Interbank 1 year - offered rate	Last vintage
Interest Rates	m	NOK: Interest rate. Interbank 5 year - middle rate	Last vintage
Interest Rates	m	NOK: Interest rate. Interbank 10 year - middle rate	Last vintage
Interest Rates	m	NOK: Interest rate. Euro rate 12 month, effective	Last vintage
Interest Rates	m	NOK: Interest Rate Swaps, Ask, 2 Year, Close	Last vintage
Interest Rates	m	NOK: Interest Rate Swaps, Ask, 3 Year, Close	Last vintage
Interest Rates	m	NOK: Interest Rate Swaps, Ask, 4 Year, Close	Last vintage
Interest Rates	m	NOK: Interest Rate Swaps, Ask, 5 Year, Close	Last vintage
Interest Rates	m	Trading partners: Interest rate 3 month	Last vintage
Commodities	m	Aluminum. USD	Last vintage
Commodities	m	Brent Blend USD/barrel	Last vintage
Surveys 1	q	Purchasing Managers Index. Outlook for next quarter	Last vintage
Labor Market 1	m	Registered unemployment rate	28.05.2009
Labor Market 1	m	Vacancies, stock	28.05.2009
Labor Market 1	m	Vacancies, supply	28.05.2009
Labor Market 1	m	Registered unemployed and government measures	28.05.2009
Labor Market 1	m	Temporary layoffs	28.05.2009
Labor Market 1	m	Government measures	28.05.2009
Labor Market 2	m	Employment. Hours	29.05.2009
Labor Market 2	m	Employment. Persons	31.07.2009
Services	m	Hotel guest nights	05.06.2009
Ind. Production	m	Total	Last vintage
Ind. Production	m	Intermediate goods	Last vintage
Ind. Production	m	Capital goods	Last vintage
Ind. Production	m	Consumer goods	Last vintage
Ind. Production	m	Durable consumer goods	Last vintage
Ind. Production	m	Non-durable consumer goods	Last vintage
Ind. Production	m	Energy goods	Last vintage
Ind. Production	m	Extraction and related services	Last vintage
Ind. Production	m	Paper and paper products	Last vintage
Ind. Production	m	Printing, reproduction	Last vintage
Ind. Production	m	Basic metals	Last vintage
Ind. Production	m	Fabricated metal products	Last vintage
Ind. Production	m	Manufacturing, mining and quarrying	Last vintage
Ind. Production	m	Manufacturing	Last vintage
Ind. Production	m	Mining and quarrying	Last vintage
Ind. Production	m	Food, beverage and tobacco	Last vintage
Ind. Production	m	Textiles, wearing apparel, leather	Last vintage
Ind. Production	m	Refined petro., chemicals, pharmaceuticals	Last vintage
Ind. Production	m	Basic chemicals	Last vintage

Block Names	Freq	Description	Start Vintage
CPI	m	Total	Last vintage
CPI	m	Consumer Price Index Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Food (without alcohol)	Last vintage
CPI	m	Alcohol & tobacco	Last vintage
CPI	m	Apparel	Last vintage
CPI	m	Housing	Last vintage
CPI	m	Furniture and Household article s	Last vintage
CPI	m	Health care	Last vintage
CPI	m	Transportation	Last vintage
CPI	m	Post- and telecom services	Last vintage
CPI	m	Culture and Leisure	Last vintage
CPI	m	Education	Last vintage
CPI	m	Hotel- and restaurant services	Last vintage
CPI	m	Other commodities and services	Last vintage
CPI	m	Electricity, gas and other fuels	Last vintage
CPI	m	Gas and lubricants	Last vintage
CPI	m	Consumer Price Index excluding Energy	Last vintage
CPI	m	Other Norwegian produced consumer goods	Last vintage
CPI	m	Other services	Last vintage
CPI	m	Other services with salary as dominated price factor	Last vintage
CPI	m	Other services including also other important price factors	Last vintage
CPI	m	Fish products	Last vintage
CPI	m	Rent (housing)	Last vintage
CPI	m	Imported consumer goods	Last vintage
CPI	m	Imported consumer goods with Norwegian competition	Last vintage
CPI	m	Imported consumer goods without Norwegian competition	Last vintage
CPI	m	Domestically produced goods and services	Last vintage
CPI	m	Agriculture goods	Last vintage
CPI	m	Other Norwegian produced consumer goods, (almost) not influenced by world maket price	Last vintage
CPI	m	Affected by world market due to competition from abroad	Last vintage
CPI	m	Affected by world market due to import share or commodity prices	Last vintage
CPI	m	Food (without alcohol). Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Alcohol & tobacco. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Apparel. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Housing. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Furniture and Household articles. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Health care. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Transportation. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Post- and telecom services. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Culture and Leisure. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Education. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Hotel- and restaurant services. Adjusted for Taxes and excluding Energy	Last vintage
CPI	m	Other commodities and services. Adjusted for Taxes and excluding Energy	Last vintage
CPI	q	International price impulses in foreign exchange	23.07.2009
CPI	q	Consumer Price Index, 25 Trading Partners	Last vintage
Foreign Trade	m	Exports, fish and fish products	14.05.2009
Foreign Trade	m	Exports, processed wood	14.05.2009
Foreign Trade	m	Exports, aluminum and -alloys	14.05.2009
Foreign Trade	m	Exports, natural gas	14.05.2009
Foreign Trade	m	Exports, crude oil	14.05.2009
Foreign Trade	m	Exports, total excluding ships and oil platforms	14.05.2009
Foreign Trade	m	Exports, total excluding ships, oil platforms, crude oil and natural gas	14.05.2009
Foreign Trade	m	Imports, total excluding ships and oil platforms	14.05.2009
Foreign Trade	m	Imports, total excluding ships, oil platforms and crude oil	14.05.2009
Construction	m	Building starts, number of residential buildings	14.05.2009

Block Names	**Freq**	**Description**	**Start Vintage**
Construction	m	Building starts, residential buildings, square meters	14.05.2009
Construction	m	Building starts, commercial buildings, square meters	14.05.2009
National accounts	q	GDP, Mainland Norway	08.06.2000
National accounts	q	Household and NPISHs consumption expenditure	08.06.2000
National accounts	q	Household consumption expenditure	08.06.2000
National accounts	q	Household consumption expenditure, goods	08.06.2000
National accounts	q	Household consumption expenditure, services	08.06.2000
National accounts	q	Public consumption expenditure	08.06.2000
National accounts	q	Gross fixed capital formation	08.06.2000
National accounts	q	Gross fixed capital formation, Mainland Norway	08.06.2000
National accounts	q	Gross fixed capital formation, Mainland Norway excluding general government	08.06.2000
National accounts	q	Gross fixed capital formation, Residential buildings	08.06.2000
National accounts	q	Gross fixed capital formation, public sector	08.06.2000
National accounts	q	Final demand from Mainland Norway (excl. changes i inventories)	08.06.2000
National accounts	q	Total exports	08.06.2000
National accounts	q	Exports, traditional goods	08.06.2000
National accounts	q	Total imports	08.06.2000
National accounts	q	Imports, traditional goods	08.06.2000
National accounts	q	Wages	29.05.2009
National accounts	q	Population 16 to 74	29.05.2009
National accounts	q	Total hours worked	29.05.2009
National accounts	q	Productivity. Mainland Norway	29.05.2009
National accounts	q	GDP, 25 Trading Partners	04.06.2009
National accounts	q	Gross Investment, Oil activities and Sea transport	19.05.2009
Surveys 2	q	Regional network, output prospects 6 months ahead. All sectors	Last vintage
Surveys 2	q	Regional network, output prospects 6 months ahead. Manufacturing industries	Last vintage
Surveys 2	q	Regional network, output prospects 6 months ahead. Export industries	Last vintage
Surveys 2	q	Regional network, output prospects 6 months ahead. Domestic market industries	Last vintage
Surveys 2	q	Regional network, output prospects 6 months ahead. Suppliers to petroleum industry	Last vintage
Surveys 2	q	Regional network, capacity utilization. All sectors	Last vintage
Surveys 2	q	Regional network, capacity utilization. Manufacturing industries	Last vintage
Surveys 2	q	Regional network, capacity utilization. Construction	Last vintage
Surveys 2	q	Regional network, capacity utilization. Services	Last vintage
Surveys 2	q	Regional network, capacity utilization. Retail trade	Last vintage
Surveys 2	q	Regional network, output growth past 3 months. All sectors	Last vintage
Surveys 2	q	Regional network, employment growth past 3 months. All sectors	Last vintage
Retail	m	Retail sales, excluding cars and gas	Last vintage
Surveys 3	q	BTS, Manufacturing. Production, change from previous quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Employment, change from previous quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. New orders in total, change from previous quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. New orders from home markets, change from previous quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. New orders from export markets, change from previous quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Prices on products at home markets, change from previous quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Prices on products at export markets, change from previous quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Production, expected change in next quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Production of intermediate goods, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Production of capital goods, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Production of consumer goods, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Employment, expected change in next quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Employment in intermediate goods sector, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Employment in capital goods sector, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Employment in consumer goods sector, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. New orders in total, expected change in next quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. New orders from home markets, expected change in next quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Prices on products at home markets, expected change in next quarter	28.07.2009

Block Names	Freq	Description	Start Vintage
Surveys 3	q	BTS, Manufacturing. General judgement of outlook in next quarter.	28.04.1993
Surveys 3	q	BTS, Manufacturing. Expected capacity utilization at end of quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Average capacity utilization, change from previous quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Average capacity utilization, expected change in next quarter	28.07.2009
Surveys 3	q	BTS, Manufacturing. Indicator of resource shortage	28.07.2009
Surveys 3	q	BTS. Industrial confidence indicator	28.04.1993
Surveys 3	q	BTS, Manufacturing. Existing plans for investment at end of quarter	28.04.1993
Surveys 3	q	BTS, Manufacturing. Stock of orders vs production expected at end of quarter	28.07.2009
Surveys 4	q	TNS Gallup. Consumer confidence, overall indicator	Last vintage
Surveys 4	q	TNS Gallup. Consumer confidence, assessment of country's economy last year	Last vintage
Surveys 4	q	TNS Gallup. Consumer confidence, buying plans	Last vintage
Surveys 4	q	TNS Gallup. Consumer confidence, assessment of own economy last year	Last vintage
Orders 1	q	Manufacturing. Supply of orders	14.08.2009
Orders 1	q	Manufacturing. Stock of orders	14.08.2009
Orders 2	q	Construction. Stock of orders. All	20.11.2009
Orders 2	q	Construction. Stock of orders. Residential buildings	20.11.2009
Orders 2	q	Construction. Stock of orders. Commercial buildings	20.11.2009
Orders 2	q	Construction. Stock of orders. Civil engineering works	20.11.2009
Orders 2	q	Construction. Supply of orders. All	20.11.2009
Orders 2	q	Construction. Supply of orders. Buildings	20.11.2009
Orders 2	q	Construction. Supply of orders. Residential buildings	20.11.2009
Orders 2	q	Construction. Supply of orders. Commercial buildings	20.11.2009
Orders 2	q	Construction. Supply of orders. Civil engineering works	20.11.2009
Credit and Money	m	Credit (C1) to general public	03.06.2009
Credit and Money	m	Credit (C2) to non-financial enterprizes	03.06.2009
Credit and Money	m	Credit (C2) to households	03.06.2009
Credit and Money	m	Credit (C2) to general public	03.06.2009
Credit and Money	q	Money Supply (M2), non-financial sector	03.08.2009
Credit and Money	q	Money Supply (M1), all sectors	03.08.2009

B Models in SAM

B.1 Vector Autoregressive (VAR) models

Classical VARs VAR models are commonly used for macroeconomic forecasting. Assume we have the following general model

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t, \qquad \epsilon_t \sim N(0, \Sigma_\epsilon), \tag{B.1}$$

where $Y_t = (y_{1,t}, y_{2,t})$. We include four different quarterly VAR models, where $y_{1,t}$ always denotes GDP when forecasting GDP and CPIATE when forecasting CPIATE, while $y_{2,t}$ is either empty (AR model) or contains CPIATE inflation (or GDP) and/or the three-month money market rate. For each of the four models we consider three different transformations of the data; first differences, double differences and detrended (using an exponential smoother). As VARs may be prone to instabilities, both full-sample and two different rolling-sample VARs are estimated. The two rolling-samples are 20 and 30 quarters long, respectively. Finally, we let the lag length p vary from 1 to 4 lags. In total, we consider 36 AR models, 72 bivariate VARs and 36 trivariate VARs. We refer to these models as the *VARs*. When forecasting CPIATE, we also include 12 AR models using monthly data.

Several papers, such as Evans (2005), Giannone et al. (2008) and Aruoba et al. (2009), show that accounting for the timeliness of data is crucial for nowcasting accuracy. To take into account the flow of data releases is therefore essential to our analysis. In the case of the VARs, both quarterly CPIATE inflation and quarterly averages of the money market rate are available prior to the release of Mainland GDP. The models are put into a state space form, and hence Kalman filter techniques can easily be applied to deal with missing data, i.e. the unbalanced data problem. This is also a key aspect when constructing the predictive densities from the different models. The forecast uncertainties are obtained through simulations, where the final densities are derived using kernel smoothing techniques. By applying the Kalman filter we can obtain conditional forecasts when the data set is unbalanced. More precisely, we use the smoothed covariance matrix of the predictors, which will resemble the mean squared error (MSE) matrix of the system, and draw from the normal distribution to obtain simulated forecasts for each horizon.[17] This is explained in more detail in section B.3.

Bayesian VARs Bayesian methods have proven useful in the estimation of VARs. In Bayesian analysis the econometrician has to specify prior beliefs about the parameters. The prior beliefs are then combined with the data in the VAR to form a posterior view of the parameters. We apply three different assumptions on the priors - uninformative, and two conjugate Normal-inverted Wishart priors (with few or many degrees of freedom). See Kadiyala and Karlsson (1997), Banbura *et al.* (2008) and Koop and Korobilis (2009) for more information on priors.

Error correction model (Emod) We estimate a vector equlibrium correction model using 13 macro variables of the Norwegian economy.[18] In addition to CPIATE, GDP and other domestic variables, foreign prices and interest rates and the oil price are included as conditional variables. Forecasts of conditional variables are projected with AR(2) processes. The model is included in the VAR class.

Dynamic Stochastic General Equilibrium (DSGE) Model The DSGE model is based on a standard New Keynesian small open economy model. A version adapted to the Norwegian economy is documented in Brubakk et al. (2006). The DSGE model is estimated using Bayesian maximum likelihood on seasonal adjusted data for mainland GDP growth, consumption growth, investment growth, export growth, employment, inflation (CPIATE), imported inflation, real wage growth, the real exchange rate (I44) and the nominal interest rate. The steady-state levels are equal to recursively updated means of the variables. The model is included in the VAR class.

[17] See for example Lütkepohl (2005).

[18] The model is documented in Akram (2008).

B.2 Leading indicator models

Bivariate classical VARs We include in total 64 leading indicators when forecasting Mainland GDP and three indicators when forecasting CPIATE (using monthly data). For each indicator we construct a bivariate VAR as described in equation B.1, where $y_{2,t}$ now will denote the leading indicator. The models are only estimated recursively. Rolling-sample indicator models may be incorporated in a future version of SAM. Most indicators are published on a quarterly frequency, while others are available at a monthly frequency. Hence, we need to bridge the monthly indicators with quarterly GDP when the purpose is to forecast GDP. This is done by constructing quarterly averages of the monthly series. If a monthly series only contains one or two months of a quarter, we simply construct the average of the one or two observations from the quarter of interest.[19] The unbalanced data problem and the construction of the predictive densities are then solved in the same way as described for the VARs above.

One-equation regression models (bridge equation) The monthly indicator models forecast GDP using several monthly indicators (that are averaged up to a quarterly frequency) as regressors. The models are estimated using OLS (ordinary least squares). In order to forecast GDP Mainland-Norway, the explanatory variables in the indicator models are projected using AR processes. The explanatory variables in the three indicator models are: manufacturing production, employment, unemployment, retail sales, hotel sleepovers and building startups. The models are included in the Indicator class.

Disaggregated model for CPIATE We first forecast the main components of CPI with AR-models. The models for the components are estimated on monthly, unadjusted data from 1991. Forecasts from each component are weighted using the consumer weights in CPI to form composite forecasts for seasonally adjusted CPIATE. The forecasts are converted to quarterly frequencies.

[19] See Baffigi et al. (2004) and Angelini et al. (2011) for a more detailed discussion of alternative bridge equations.

B.3 Factor Models

The objective of factor models is to summarize the information contained in large datasets by reducing their dimension, ie reducing the parameter space. The model that we use is an approximate monthly dynamic factor model similar to Giannone et al. (2008). This is a model that account for the unbalanced data problem. Assume we have a vector of n stationary monthly variables $X_t = \begin{pmatrix} x_{1t}, & \dots, & x_{nt} \end{pmatrix}'$, $t = 1, ..., T$, which have been standardized to have zero mean and variance equal to one. The model is given by the following two equations.

$$X_t = \chi_t + \xi_t = \Lambda F_t + \xi_t, \qquad \xi_t \sim N(0, \Sigma_\xi) \tag{B.2}$$

$$F_t = \sum_{i=1}^{p} A_i F_{t-i} + B u_t, \qquad u_t \sim N(0, I_u) \tag{B.3}$$

Equation B.2 relates the monthly time series X_t to a common (unobserved) component χ_t plus an idiosyncratic component $\xi_t = (\xi_{1,t}, \dots, \xi_{n,t})'$. The common component can be decomposed in an $r \times 1$ vector of latent factors $F_t = (f_{1,t}, \dots, f_{r,t})'$ times a $n \times r$ matrix of factor loadings Λ. The idiosyncratic component is assumed to be multivariate white noise. Equation B.3 describes the law of motion for the latent factors. The factors are driven by q-dimensional standardized white noise u_t, where B is an $r \times q$ matrix, where $q \leq r$. Finally, $A_1, \dots, A_p$ are $r \times r$ matrices of parameters.

Our task is to forecast quarterly growth in GDP or four-quarter growth in CPIATE (y_t^Q). We need to build a bridge between the monthly explanatory variables and the quarterly series that we want to forecast. All monthly variables are transformed to ensure that the corresponding quarterly quantities are given by $x_{i,t}^Q \sim (x_{i,t} + x_{i,t-1} + x_{i,t-2})$ measured at the last month of each quarter, i.e. $t = 3k$ and $k = 1, \dots, T/3$. This implies that series in differences enter the factor model in terms of three-month changes. Defining the quarterly factors as $F_t^Q = (F_t + F_{t-1} + F_{t-2})$, the factors-based bridge equation follows:

$$y_t^Q = \alpha + \beta' \widehat{F}_t^Q + e_t, \qquad e_t \sim N(0, \Sigma_e) \tag{B.4}$$

where β is an $r \times 1$ vector of parameters.

The model is estimated in a two-step procedure using principal components and the Kalman filter. The unbalanced part of the data set can be incorporated through the use of the Kalman filter, where missing observations are interpreted to have an infinitely large noise to signal ratio. For more details about this, see Giannone et al. (2008).

The simulated forecasts from the dynamic factor model are derived using a small modification of the technique described for the VARs above. The factors are derived and forecasted using the Kalman filter, while the forecasting equation itself is direct, and conditional on the factors. Uncertainty in the factor forecasts are taken into account in the forecasting equation. The factor uncertainty is drawn from the normal distribution, using the smoothed covariance matrix from the Kalman Filter, in the same manner as described above. The uncertainty in the forecasting equation is derived through drawing a random shock from the normal distribution of past residuals.

C Table and figures

Table C.2. *Structure of data releases, number of models updated for each block of information and model class, and forecast horizon, from the start of the quarter until the first estimate of GDP/CPIATE is released*

Blocks of information			Mainland GDP					CPIATE				
Block	Data	Time	Horizon	VAR	Indicator	Factor	SAM	Horizon	VAR	Indicator	Factor	SAM
1	Credit		2					2				
2	Financials	m1	2	5	4	4	13	2	5	1	4	10
3	Interest rate		2	77		4	81	2	77	1	4	82
4	Commodoties		2	1		4	5	2	1		4	5
5	PMI		2		1		1	2				0
6	Labor		2	1	6	5	12	2	1	1	5	7
7	Services		2		3	4	7	2			4	4
8	Ind. Production		2		3	4	7	2			4	4
9	CPI		2	77	1	5	83	1	89	3	5	97
10	Foreign trade		2			4	4	1			4	4
11	Construction		2		3	4	7	1			4	4
12	BTS		2		33	1	34	1			1	1
13	Retail		2		4	4	8	1			4	4
14	Credit		2		6	5	11	1			5	5
15	Financials	m2	2			4	4	1		1	4	5
16	Interest rate		2			4	4	1		1	4	5
17	Commodoties		2			4	4	1			4	4
18	Labor		2		3	4	7	1		1	4	5
19	Services		2		3	4	7	1			4	4
20	Ind. Production		2		3	4	7	1			4	4
21	Orders industry		2		3	1	4	1			1	1
22	CPI		2			4	4	1	12	3	4	19
23	Foreign trade		2			4	4	1			4	4
24	Construction		2			4	4	1			4	4
25	GDP		1	149	60	1	210	1	149		1	150
26	Regional network		1		12		12	1				0
27	Orders construction		1		10	1	11	1			1	1
28	Retail		1		3	4	7	1			4	4
29	Credit		1			4	4	1			4	4
30	Financials	m3	1			4	4	1		1	4	5
31	Interest rate		1			4	4	1		1	4	5
32	Commodoties		1			4	4	1			4	4
33	Labor		1		7	4	11	1		1	4	5
34	Services		1		3	4	7	1			4	4
35	Ind. Production		1		3	4	7	1			4	4
36	Consumer confidence		1		6	1	7	1			1	1
37	CPI		1			4	4	1	12	3	4	19
38	Foreign trade		1			4	4	1			4	4
39	Construction		1		3	4	7	1			4	4
40	Retail		1		3	4	7	1			4	4
41	Credit		1			4	4	1			4	4
42	Financials	m4	1	5	4	4	13	1	5	1	4	10
43	Interest rate		1	5		4	9	1	5	1	4	10
44	Commodoties		1	1		4	5	1	1		4	5
45	PMI		1		1		1	1				0
46	Labor		1	1	6	5	12	1	1	1	5	7
47	Services		1		3	4	7	1			4	4
48	Ind. Production		1		3	4	7	1			4	4
49	CPI		1	77	1	5	83					
50	Foreign trade		1			4	4					
51	Construction		1			4	4					
52	BTS		1		33	1	34					
53	Retail		1		4	4	8					
54	Credit		1		6	5	11					
55	Financials	m5	1			4	4					
56	Interest rate		1			4	4					
57	Commodoties		1			4	4					
58	Labor		1		3	4	7					
59	Services		1		3	4	7					
60	Ind. Production		1		3	4	7					
61	Orders industry		1		3	1	4					
62	CPI		1			4	4					
63	Foreign trade		1			4	4					
64	Construction		1		3	4	7					

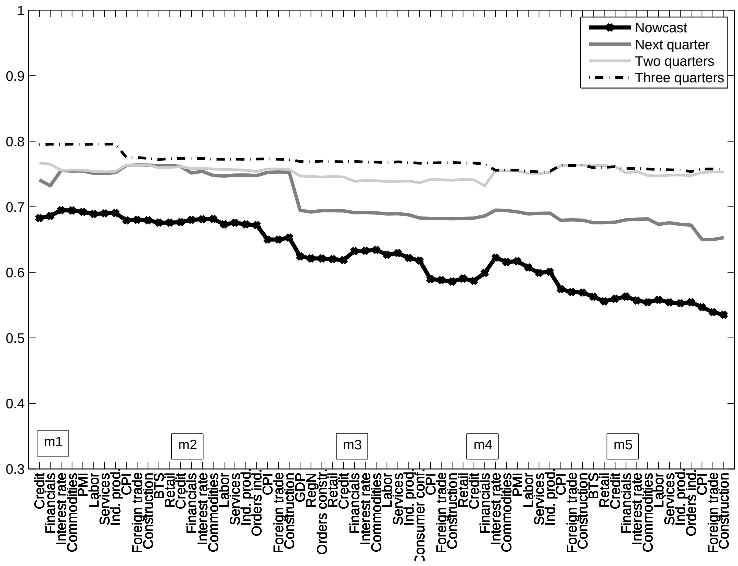


Figure C.1. *RMSE for forecasts of quarterly growth in Mainland GDP after different block releases and for different horizons. Evaluated against 5th release of outcomes.*

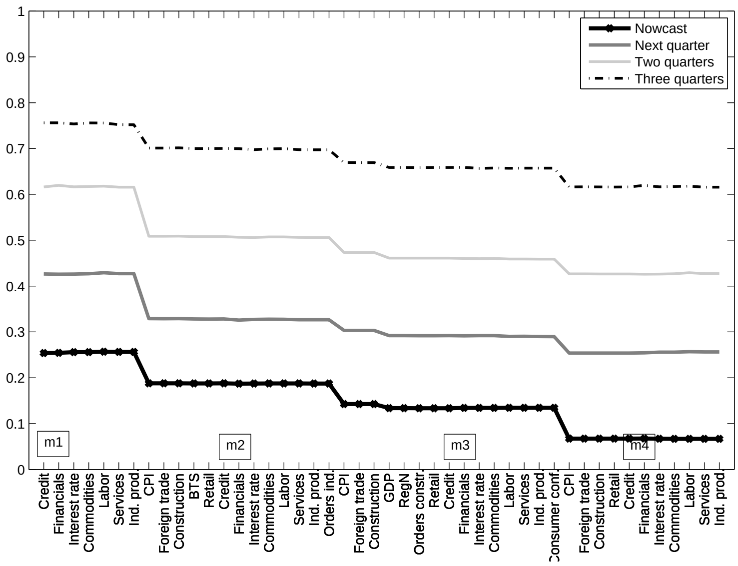


Figure C.2. *RMSE for forecasts of four-quarter growth in CPIATE after different block releases and for different horizons.*

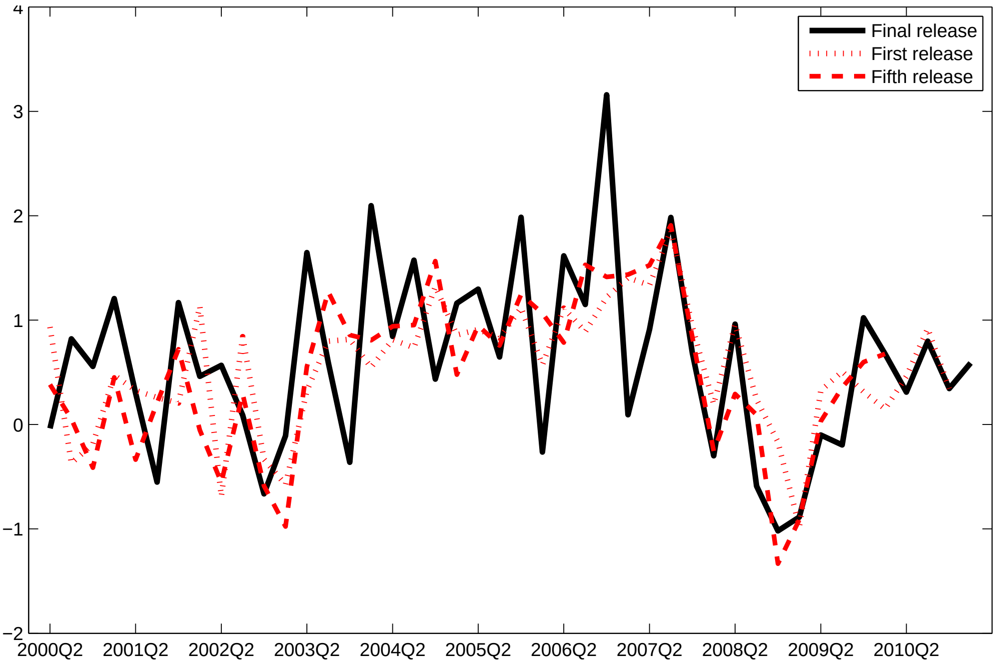
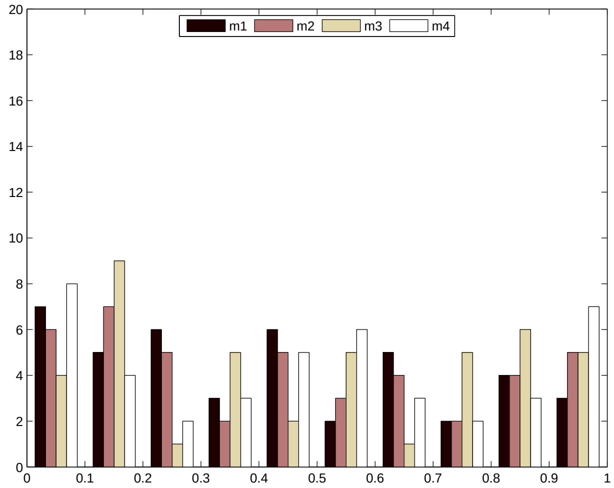
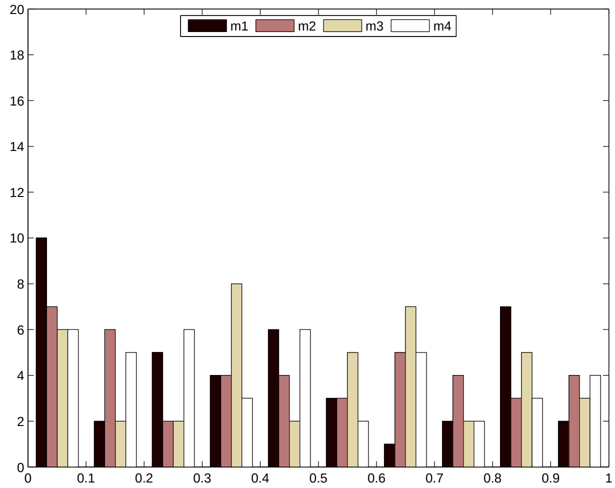


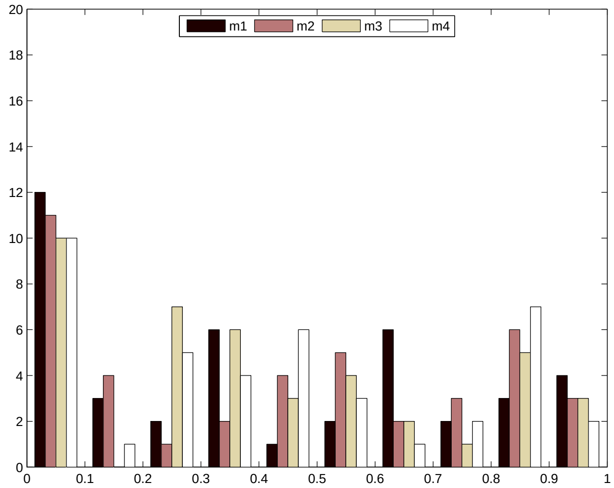
Figure C.3. *First release, fifth release and final vintage of seasonally adjusted Norwegian Mainland-GDP. Quarterly growth rates. Per cent*



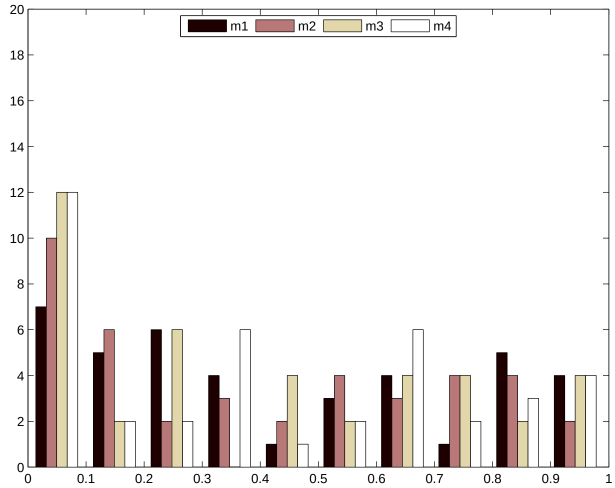
(a) Nowcast



(b) Next quarter

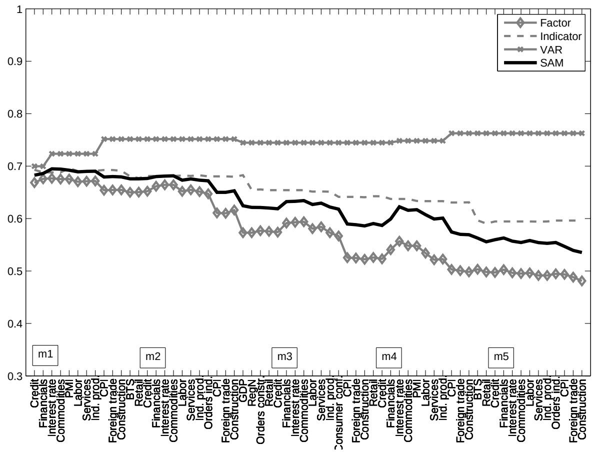


(c) Two quarters

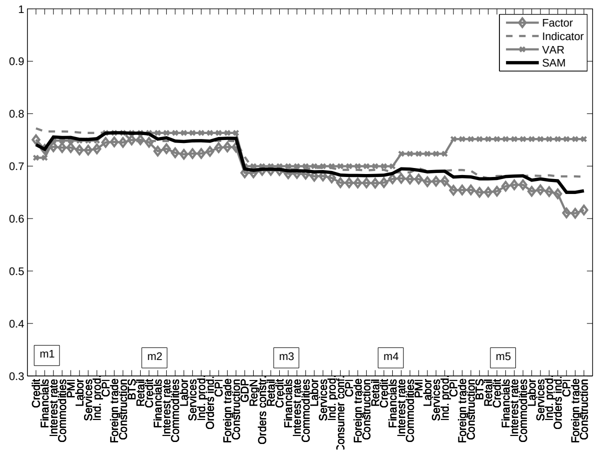


(d) Three quarters

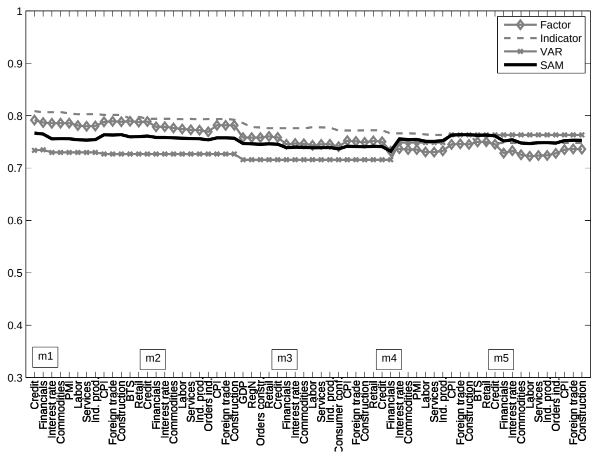
Figure C.4. *Probability integral transforms (pits) at different points in time. Log-score weighted forecasts of four-quarter growth in CPIATE. The pits of a forecasting model should have a standard uniform distribution if the model is correctly specified.*



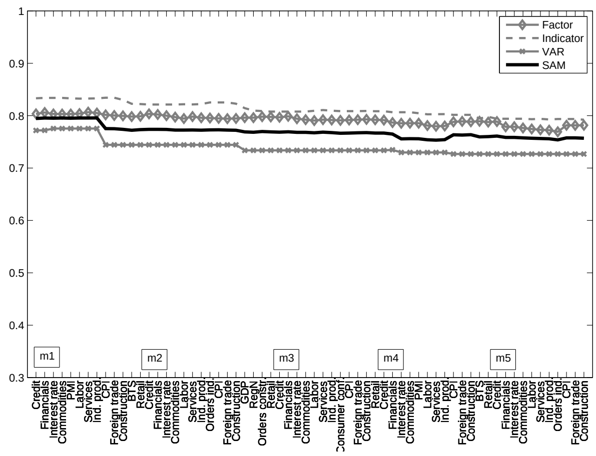
(a) Nowcast



(b) Next quarter

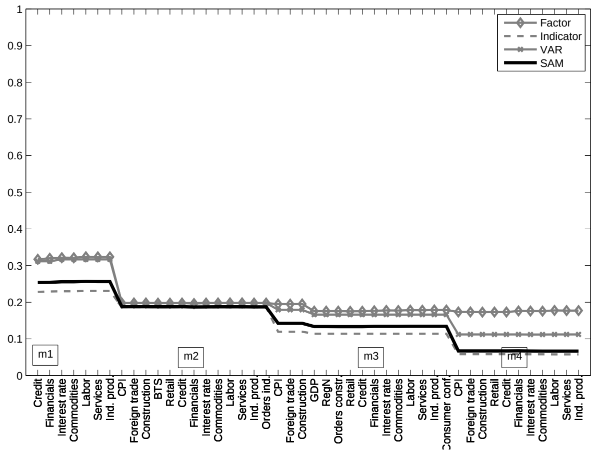


(c) Two quarters

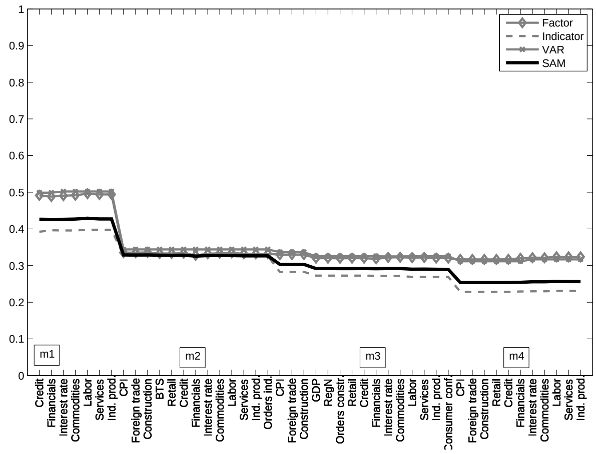


(d) Three quarters

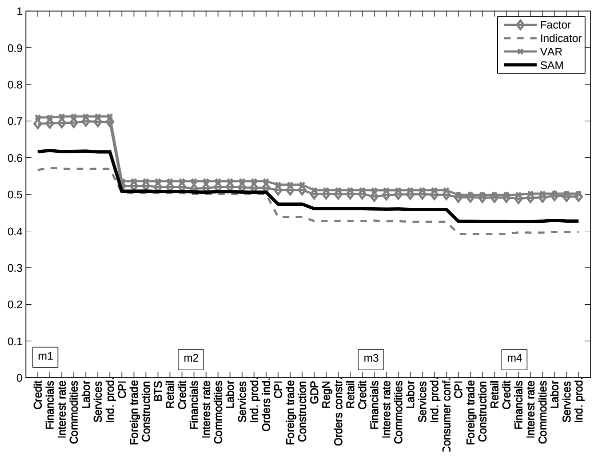
Figure C.5. *RMSE for forecasts made after different block releases for model classes and SAM for different horizons. Forecasts of quarterly growth in Mainland GDP. Evaluated against 5th release of Mainland GDP*



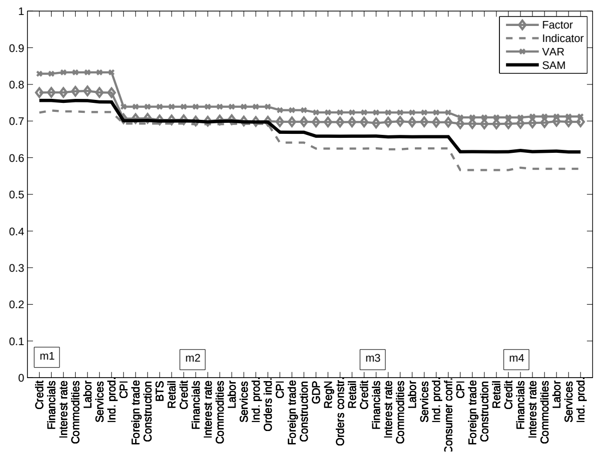
(a) Nowcast



(b) Next quarter

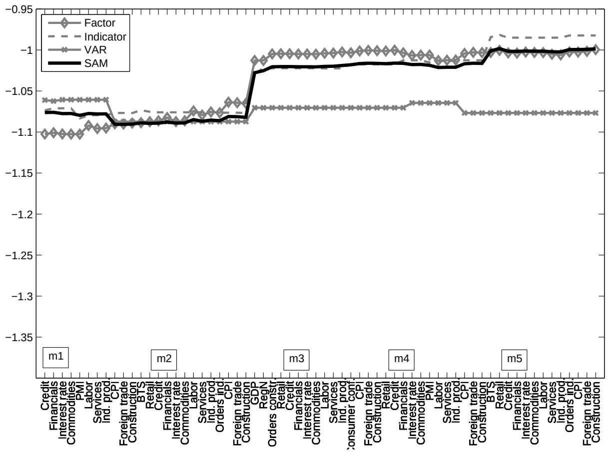


(c) Two quarters

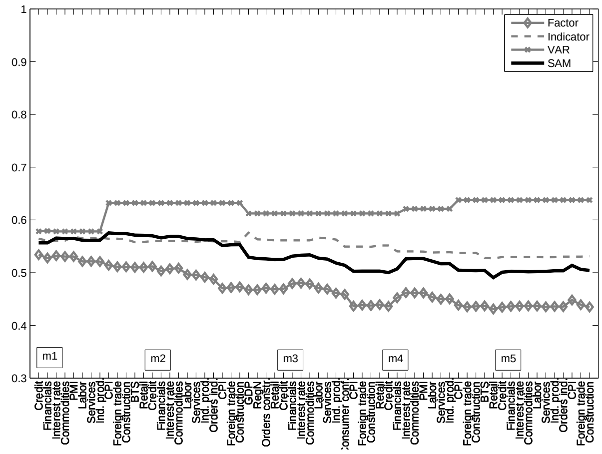


(d) Three quarters

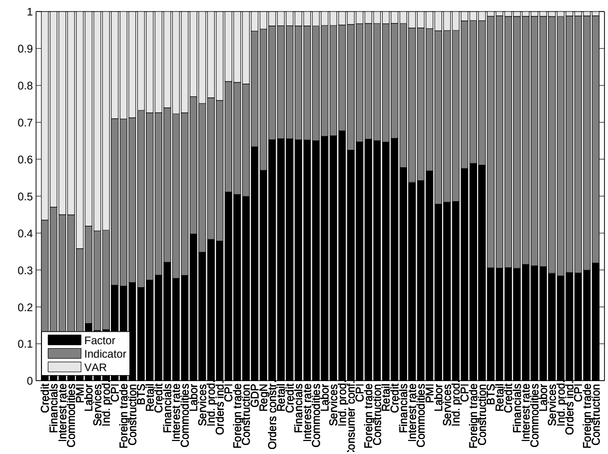
Figure C.6. *RMSE for forecasts made after different block releases for model classes and SAM for different horizons. Forecasts for four-quarter growth in CPIATE*



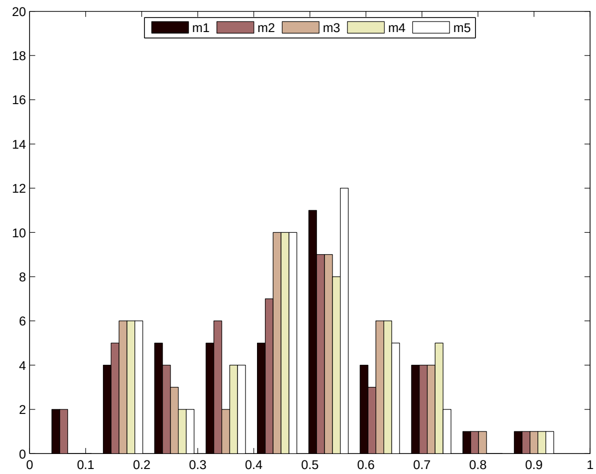
(a) Average log score



(b) RMSE

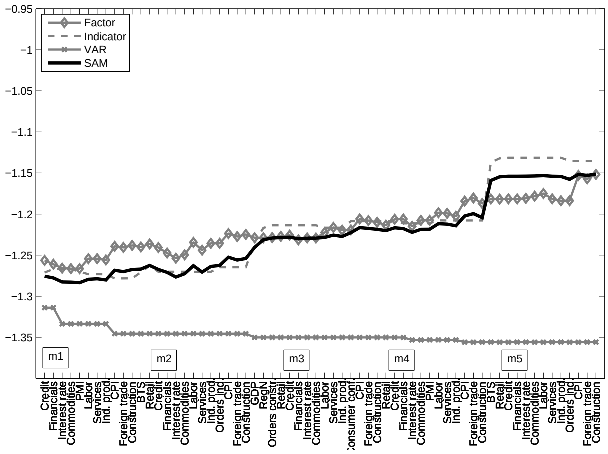


(c) Weights

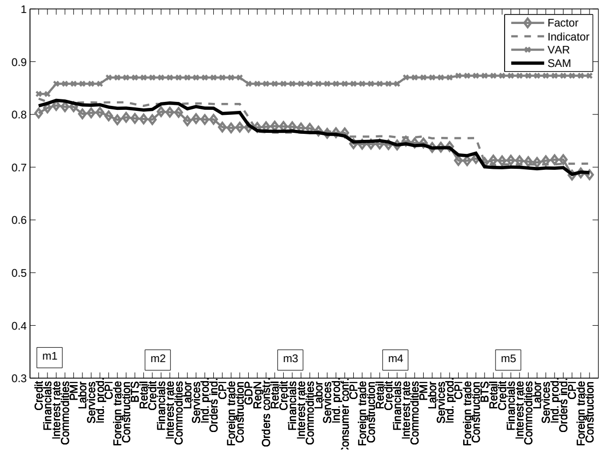


(d) Pits

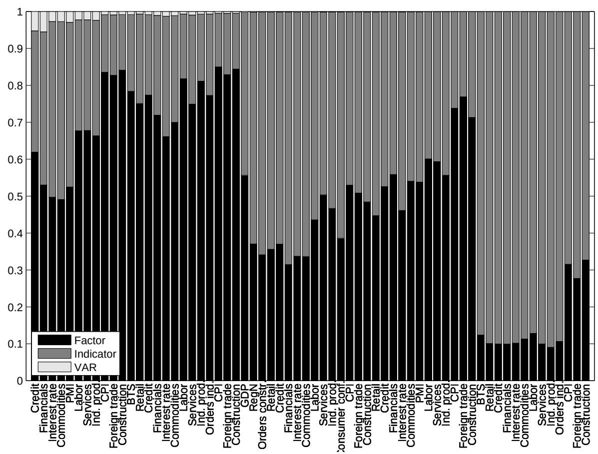
Figure C.7. Robustness of results when evaluated against 1st release of data. Forecasts for quarterly growth in Mainland GDP



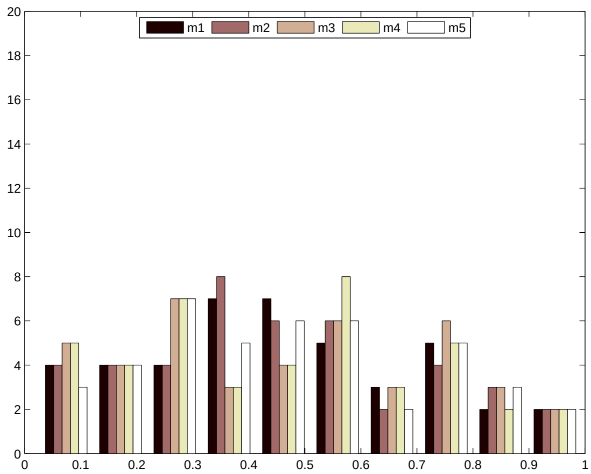
(a) Average log score



(b) RMSE

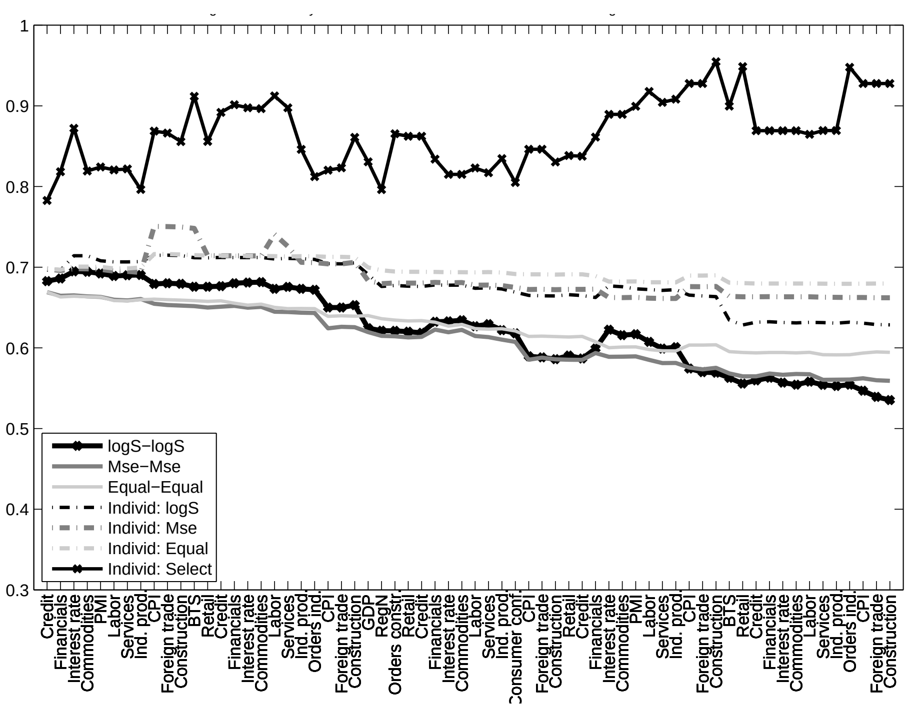


(c) Weights

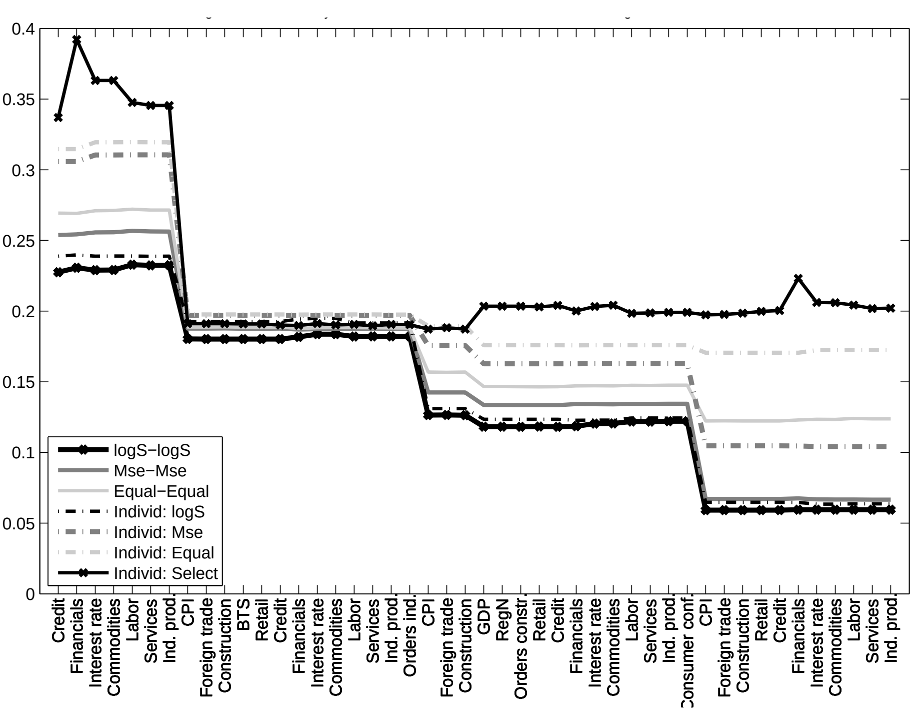


(d) Pits

Figure C.8. Robustness of results when evaluated against final vintages of data. Forecasts for quarterly growth in Mainland GDP



(a) GDP Mainland-Norway



(b) CPI-ATE

Figure C.9. *Comparing different weighting schemes. RMSEs for forecasts after different block releases. Forecasts of quarterly growth in Mainland GDP evaluated against 5th release of outcomes, and forecasts of four-quarter growth in CPIATE.*