

No. 12 | 2012

# Staff Memo

## Evaluating real-time forecasts from Norges Bank's system for averaging models

*Anne Sofie Jore, Norges Bank Monetary Policy*

Staff Memos present reports and documentation written by staff members and affiliates of Norges Bank, the central bank of Norway. Views and conclusions expressed in Staff Memos should not be taken to represent the views of Norges Bank.

© 2012 Norges Bank

The text may be quoted or referred to, provided that due acknowledgement is given to source.

Staff Memo inneholder utredninger og dokumentasjon skrevet av Norges Banks ansatte og andre forfattere tilknyttet Norges Bank. Synspunkter og konklusjoner i arbeidene er ikke nødvendigvis representative for Norges Banks.

© 2012 Norges Bank

Det kan siteres fra eller henvises til dette arbeid, gitt at forfatter og Norges Bank oppgis som kilde.

**ISSN 1504-2596 (online only)**

**ISBN 978-82-7553-6\* - !%(online only)**

# Evaluating real-time forecasts from Norges Bank's System of Averaging Models

Anne Sofie Jore

April 10, 2012

## **Abstract**

We evaluate forecasts of GDP growth and inflation made by the system of averaging models (SAM) and compare their performance with Norges Bank's forecasts in the *Monetary Policy Report*. We assess the new version of SAM, as well as the ones that were actually used from 2008 until the end of 2010. We conclude that SAM forecasts perform quite well. In particular for inflation, the forecasting performance seems to have improved with the new version of SAM.

# 1 Introduction

Policy decisions in real time are based on assessments of the recent past and current economic conditions under a high degree of uncertainty. 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.

Short-term forecasts at central banks are typically formed by sector experts' use of econometric models and their views on economic developments. Since 2008, a System of Averaging Models (SAM) has played an important role in the short-term forecasting process in Norges Bank. SAM is documented in [Aastveit et al. \(2011\)](#).

SAM forecasts have been available since *Monetary Policy Report* (MPR) 2/2008. A period of three to four years is too short for an extensive evaluation of the SAM forecasts, but in our view it will be useful to monitor the forecasting process by evaluating and comparing forecasts regularly. This evaluation is a first step.

In this paper we evaluate forecasts from SAM since MPR 2/2008 and compare these with official forecasts from Norges Bank published in the MPR. SAM forecasts have been playing a gradually more important role as input in the MPR forecasting process, as a supplement to other models used by sector experts. Models used by sector experts are typically conditioned on deterministic forecasts of other variables. Only a few such conditional models are included in SAM. Furthermore, judgements concerning the short-term economic situation are always important - sometimes crucial.

SAM has been developed further since we first started to use the system. The current version has been used since MPR 1/2011. We use this new version and calculate forecasts from early 2004 onwards. We use real time data to compute the forecasts. The forecasts can thus be interpreted as the forecasts we would have had if SAM had existed at the time. Then we are able to evaluate SAM forecasts and compare them with MPR forecasts over a longer period. At the same time we are able to assess if the new version of SAM performs better than earlier versions. Throughout the paper, we will refer to the earlier versions of SAM as SAM-1 and the current version as SAM-2.

National accounts data are heavily revised. Statistics Norway publishes a new vintage of national accounts data every quarter. Their first estimate of a quarter is published around 7 weeks after the end of the quarter, and at the same time previous quarters are revised. The revised numbers are based on increasingly more information, and thus are closer to the “true” values. We have several possibilities when choosing which revision to use as actual data. Depending on which revision we use as actual data in the evaluation, we may in general get different results. In this evaluation we will focus on the first available estimate of growth in GDP Mainland Norway. To check how robust this approach is, we have also evaluated the forecasts using the second and the fifth releases, respectively. The results are very similar. One reason for focusing on the first releases is the limited size of the sample. Secondly, we interpret first releases to be rational forecast of the fifth releases.

The rest of the paper is organized as follows: In section 2 we evaluate forecasts for four-quarter growth in Gross Domestic Product in Mainland Norway (GDP) and inflation adjusted for tax changes and excluding energy prices (CPIATE). In section 2.1 we evaluate and compare actual real time SAM forecasts (SAM-1) with forecasts from MPR. In section 2.2 we use the new version of SAM (SAM-2) and compute forecasts from early 2004, in accordance with publications of Monetary Policy reports.<sup>1</sup> We split the sample period in two parts: The early part contains forecasts from a period when SAM did not exist, and the late part starting with MPR 2/2008 is the period when SAM forecasts were published in the Monetary Policy reports. In section 2.3 we check the robustness of the evaluation of GDP forecasts in two directions: Evaluating quarter-to-quarter growth rates instead of four-quarter rates, and using alternative vintages of historical data as actual outcomes. We conclude in section 3.

---

<sup>1</sup>We will refer to Norges Bank’s forecasts published in MPRs (or Inflation Reports before 2007) as the “MPR forecasts”. Forecasts from SAM made in connection with MPR forecasts will be referred to as “SAM forecasts in MPR”.

## 2 Evaluation of forecasts for GDP and CPI-ATE

Forecasts from SAM have been published since MPR 2/2008. In MPR 2/2008, SAM was presented in the box “SAM - System of models for short-term forecasting”. Inflation forecasts from this early version of SAM were evaluated in [Bjørnland et al. \(2012\)](#). The evaluation period was 2001 to 2009, and the main result was that SAM forecasts outperformed Norges Bank forecasts for all horizons.

Since we started to use SAM in 2008, SAM has been developed further in several steps. In the earliest version we combined the 8 best-performing point forecasts, based on the historical performance of the forecasts, see [Bjørnland et al. \(2012\)](#). The combination framework was then extended to cover more models, and the weights were calculated using the complete forecast densities (see [Bjørnland et al. \(2011\)](#) for details). Since MPR 2/2010 Norges Bank has published forecasts from SAM as fan charts, emphasising the importance of assessing the uncertainty around the central forecast. The other major change was grouping the different models in separate model classes or ensembles, and combining the forecasts in two steps. There were 16 model classes for GDP and 10 classes for CPI-ATE. In the first step we combined forecasts in each class into a single forecast density (or ensemble density). In the second step we combined the ensemble densities into a new single forecast density, or “grand ensemble density” (see [Gerdrup et al. \(2009\)](#) for details). The current (since MPR 1/2011) version of SAM is documented in [Aastveit et al. \(2011\)](#). The two-step procedure is simplified by dividing the models into only three different model classes: vector autoregressive models (essentially combinations of GDP, inflation and the interest rate), factor models (a large number of monthly and quarterly data) and finally simple “leading indicator models” covering surveys in addition to other quarterly and monthly data.

The forecast evaluation in this paper covers the period 2004 to 2011. We evaluate point forecasts of inflation and four-quarter growth rates of GDP in the mainland economy using real time data to construct forecasts for the period 2004 to 2008.

Hence, we step back in time and use only information that was available at the time.<sup>2</sup> SAM is developed since 2008 with knowledge of economic conditions prior to 2008, but this informational advantage is deemed to be not very important. Most of the models are purely statistical, with no attempt to fit any specific model specification to the data. Furthermore, the estimated conditional indicator models for GDP were developed in 2005.

The exercise is divided in two parts. In section 2.1, we evaluate the actual SAM forecasts (SAM-1) for inflation and GDP growth from MPR 2/2008 to MPR 3/2011. In this period, SAM forecasts became gradually more important as input in the forecasting process. In section 2.2, we extend the evaluation period to the beginning of 2004, using the current version of SAM (SAM-2) to calculate real-time forecasts. We compare SAM forecasts with Norges Bank forecasts published in the *Monetary Policy Report* (or *Inflation Report* (IR)) and use simple autoregressive models as benchmarks. In section 2.3 we do some robustness checks for the GDP evaluation using alternative vintages as actual historical outcomes.

## 2.1 SAM-1 and MPR forecasts from MPR 2/2008 to MPR 3/2011

### 2.1.1 GDP forecasts

In the period from MPR 2/2008 to MPR 3/2011 SAM-1 forecasts gradually played a more important role as input in the forecasting process. Chart 1 displays MPR forecasts of four-quarter growth in GDP and SAM-1 forecasts calculated in connection with the respective reports. Actual growth rates are first releases of four-quarter GDP growth.

The first SAM-1 forecasts, from June 2008 (MPR 2/2008), pointed toward continued decreasing four-quarter growth for a few quarters and then higher growth. Norges Bank forecasts in MPR show a similar development, although growth rates are lower and

---

<sup>2</sup>We have saved vintages of national accounts data since June 2002. For some variables that are not revised, like the Business Tendency Survey, we have recursively truncated the unadjusted data and seasonally adjusted them. For other explanatory variables, for instance manufacturing production and retail sales, vintages are saved since May 2009.

closer to actual outcomes. MPR forecasts from October 2008 (MPR 3/2008) anticipate partly the continuing downturn, taking into account information not yet reflected in most short-term statistics. Since this information was not available to SAM, SAM-1 forecasts are very far away from realised outcomes. This is a good illustration of the importance of using judgement in short term forecasting. In October 2008, sentiments were very negative after the bankruptcy of Lehman Brothers in September. Stock prices plummeted, the exchange rate depreciated and short-term interest rates fell. These developments were only partly reflected in the data, and forecasts clearly had to be based mainly on judgement. MPR forecasts in March, from MPR 1/2009, turned out to be fairly accurate, while SAM-1 forecasts were still far too high. Since MPR 2/2009, MPR and SAM-1 forecasts have told the same story and forecasts have been quite close to actual outcomes.

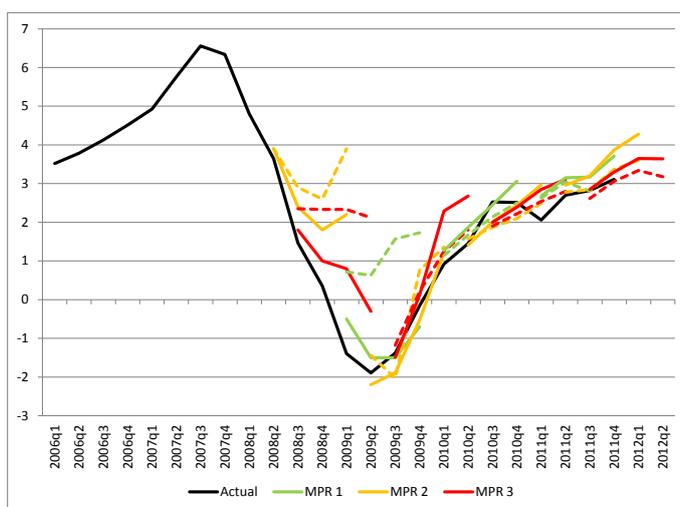


Figure 1. *Seasonally adjusted Mainland GDP. Year-on-year first release growth rates, percent. Real-time forecasts from MPR (solid lines) and SAM-1 (dashed lines). Green, yellow and red lines represent MPRs from late March (MPR 1), late June (MPR 2) and late October (MPR 3), respectively. First sequence of forecasts from MPR 2/2008.*

Forecast performance is measured by calculating the Root Mean Square Forecast Error (RMSE) for each horizon (see Table 1). As can be inferred from the early forecasts in Chart 1, average forecast errors in this period are larger for SAM-1 forecasts than for MPR forecasts.

Table 1. RMSE. Real-time forecasts for GDP, four-quarter growth. Forecasts from MPR 2/08 to MPR 3/11. Figures in bold indicate the best-performing forecast.

	MPR 2/08 to MPR 3/11	
	MPR	SAM-1
<b>h=1</b>	<b>0.42</b>	0.76
<b>h=2</b>	<b>0.49</b>	1.11
<b>h=3</b>	<b>1.01</b>	1.71
<b>h=4</b>	<b>1.45</b>	2.31

In Chart 2 we have calculated Root Mean Square Forecast Errors (RMSE)<sup>3</sup> for the whole forecast horizon (four quarters) for each *Report*. As we already know from Chart 1, forecast accuracy improves markedly from MPR 2/2009 onwards. When forecast accuracy is relatively high, the performances of SAM-1 and MPR are quite similar. Forecasts from SAM were clearly more accurate than PPR forecasts in PPR 3/09.

To sum up, this exercise has demonstrated that a broad approach is useful when forecasting short-term developments. Sometimes it is obvious that “technical” forecasts fail to take account of unexpected events, and judgements are crucial. At other times, “technical” forecasts track actual developments well and could play an important role as part of the forecasting process in Norges Bank.

---

<sup>3</sup>RMSE per report is computed as

$$\sqrt{\frac{1}{H} \sum_{h=1}^H (Y_h - F_h)^2},$$

where  $Y$  is the actual value and  $F$  is the forecast, both at horizon  $h$ , and  $H = 4$ .

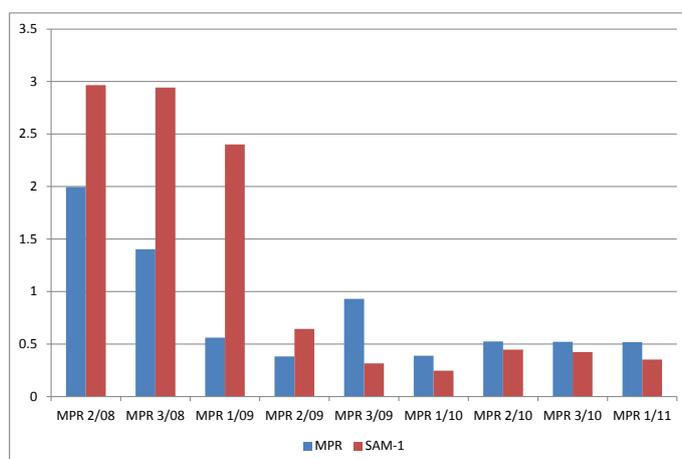


Figure 2. *RMSEs per report, horizons 1 to 4. Seasonally adjusted Mainland GDP, four-quarter growth. Real-time forecasts. MPR 2/2008 to MPR 1/2011*

### 2.1.2 Inflation forecasts

Inflation forecasts are shown in Chart 3. SAM-1 forecasts in MPR 2/2008 were too low two steps ahead, but remarkably close for the longer horizons. SAM-1 forecasts as well as MPR forecasts in later reports turned out to be fairly close to actual developments, except for the last report in 2009 and the first two reports in 2010. In these reports forecasts pointed towards a leveling out of inflation in 2010, but inflation continued to decrease. In the evaluation of the projections for 2010 in MPR 1/11, the main reasons for the lower-than-expected inflation were ascribed to weaker external price impulses and a stronger exchange rate than expected.

Forecast accuracy for each horizon is summarised in Table 2. For all horizons, MPR forecasts and SAM-1 forecasts are very similar in terms of forecast accuracy. This finding is in contrast to the results in Bjørnland et al. (2012), where it was found that SAM forecasts outperformed MPR forecasts for all horizons. The exercise in Bjørnland et al. (2012) was undertaken before we started to use SAM in the MPR forecasting process, while the exercise in this paper is an evaluation of actual forecasts. Forecasts from

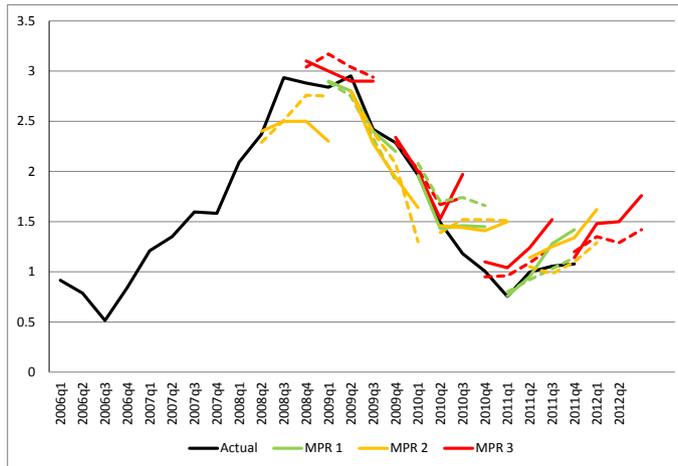


Figure 3. *Seasonally adjusted CPIATE. Four-quarter growth rates, per cent. Real-time forecasts from MPR (solid lines) and SAM-1 (dashed lines). Green, yellow and red lines represent MPRs from late March (MPR 1), late June (MPR 2) and late October (MPR 3), respectively. First sequence of forecasts from MPR 2/2008.*

SAM-1 were gradually playing a more important role in the MPR forecasting process, and thus we have reason to believe that MPR forecasts benefitted from incorporating SAM-1 in the forecasting process. This is supported by comparing RMSEs in Bjørnland et al. (2012) with RMSEs in table 2. Forecast accuracy for SAM forecasts are roughly in line in the two studies, while RMSEs for MPR forecasts are clearly smaller in the new study.

Looking at RMSEs per report in Chart 4, we observe that average SAM-1 forecasts over 4 horizons are increasingly accurate towards the end of the evaluation period. The accuracy of MPR forecasts are also improving, but to a lesser extent the SAM-1 forecasts.

Table 2. RMSE. Real-time forecasts for CPI-ATE, four-quarter growth. Forecasts from MPR 2/08 to MPR 3/11. Figures in bold indicate the best-performing forecast.

	MPR 2/08 to MPR 3/11	
	MPR	SAM-1
<b>h=1</b>	0.10	<b>0.10</b>
<b>h=2</b>	<b>0.21</b>	0.23
<b>h=3</b>	<b>0.26</b>	0.26
<b>h=4</b>	0.51	<b>0.50</b>

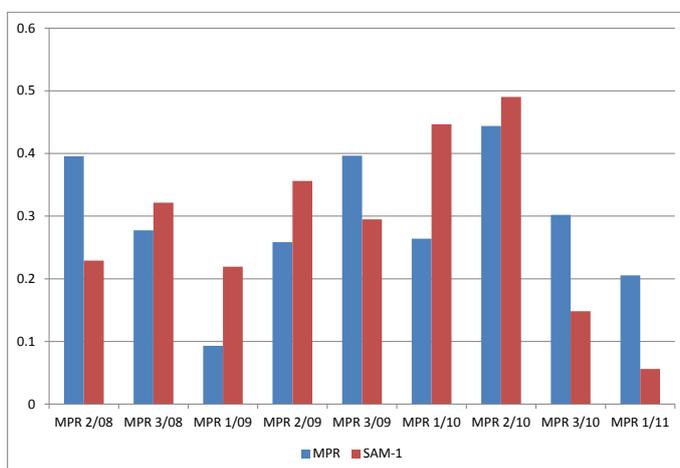


Figure 4. *RMSE per report, horizons 1 to 4. CPI-ATE, Four-quarter growth rates. Real-time forecasts. MPR2/2008 to MPR 1/2011.*

## 2.2 Forecasts from MPR 1/2004 to MPR 3/2011

To enable a more extensive evaluation of the new version of SAM, we have calculated SAM-2 forecasts for the period covering MPR 1/2004 (or IR 1/2004), to MPR 3/2010<sup>4</sup>, using real-time data. In section 2.2.1 we evaluate GDP growth forecasts and in section 2.2.2 inflation forecasts.

<sup>4</sup>The new version was first used in MPR 1/2011.

## 2.2.1 GDP forecasts

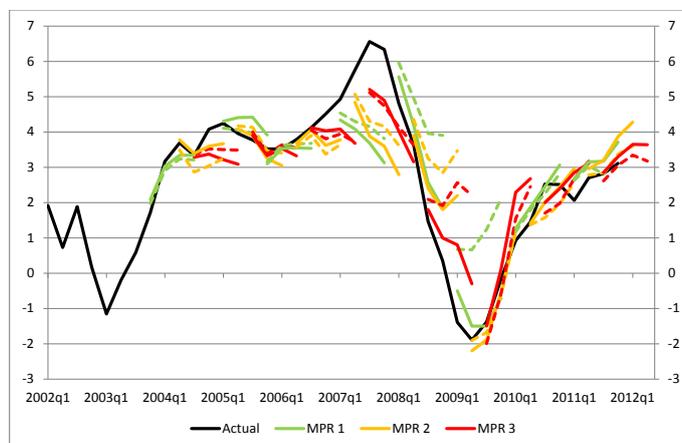


Figure 5. *Seasonally adjusted Mainland GDP. Four-quarter first release growth rates, per cent. Forecasts from MPR (solid lines) and SAM-2 (dashed lines). Green, yellow and red lines represent MPRs from late March (MPR 1), late June (MPR 2) and late October (MPR 3), respectively. First sequence of forecasts from MPR 1/2004.*

Prior to MPR 2/2008, short-term forecasts of GDP growth 3 and 4 quarters ahead were not regularly published in the *Monetary Policy/Inflation Report*. In some reports forecasts for two quarters were published. For these reports, forecasts 3 and 4 quarters ahead are calculated on the basis of published annual forecasts. In other reports, Norges Bank only published forecasts for the output gap, and growth rates are calculated by using output gaps and Norges Bank's estimates at the time of potential growth of GDP Mainland Norway.

Forecasts for four-quarter growth in GDP Mainland Norway from MPR as well as SAM-2 seem quite accurate in the first couple of years, see Chart 5.<sup>5</sup> When growth started to increase in 2006, however, forecasts stayed around the same level as the years before. Forecasts from 2007 were pointing to decreasing growth, while growth turned out

<sup>5</sup>Actual growth rates are first releases of four-quarter GDP growth.

to be accelerating. The slowdown did eventually happen, but not until winter 2007/2008. The “technical” and judgement free forecasts from SAM-2 during this period are in line with Norges Bank’s forecasts in the *Monetary Policy Report*. This may indicate that the information we had access to at the time did not point towards the strong growth that actually occurred, and that the economy was hit by an unexpected shock. In MPR 1/2008 projections for 2007 were evaluated. The record high economic growth in 2007 were ascribed to unexpectedly high growth in labour supply due to immigration and stronger than projected growth in the world economy.

In MPR 1/2008, four-quarter growth rates were expected to decrease from more than 5 per cent in the first quarter of 2008 to less than 2 per cent in the last quarter of 2008. Even this pronounced deceleration of the forecasted growth did not quite match the actual rapid decrease of the growth rates. Forecasts from SAM-2 pointed to an even less pronounced downturn.

Forecasts from MPR 2/2008 onwards are described in section 2.1.1. Although the SAM forecasts in Charts 1 and 5 are calculated with different versions of SAM for most of the period, the forecasts are quite similar and they tell the same story.

Table 3. RMSE. Forecasts for GDP, four-quarter growth. Forecasts from IR 1/2004 to MPR report 3/2011. Forecasts in IR/MPR, forecasts with SAM-2 and forecasts with an AR(4) model. Numbers in bold indicate the best-performing forecast.

	MPR 1/04 to MPR 1/08			MPR 2/08 to MPR 3/11			MPR 1/04 to MPR 3/11		
	MPR	SAM-2	AR(4)	MPR	SAM-2	AR(4)	MPR	SAM-2	AR(4)
<b>h=1</b>	<b>0.55</b>	0.59	0.74	<b>0.42</b>	0.78	1.05	<b>0.50</b>	0.68	0.89
<b>h=2</b>	1.02	<b>0.99</b>	1.12	<b>0.49</b>	1.12	1.48	<b>0.82</b>	1.05	1.29
<b>h=3</b>	<b>1.27</b>	1.28	1.45	<b>1.01</b>	1.74	2.14	<b>1.17</b>	1.50	1.78
<b>h=4</b>	<b>1.37</b>	1.46	1.71	<b>1.45</b>	2.29	2.72	<b>1.41</b>	1.85	2.18

In table 3 we have divided the sample period in two: MPR 1/2004 to MPR 1/2008 and MPR 2/2008 to MPR 3/2011. In the latest period, SAM forecasts were published in the MPR. As a reference we also show RMSEs from an autoregressive model with 4 lags

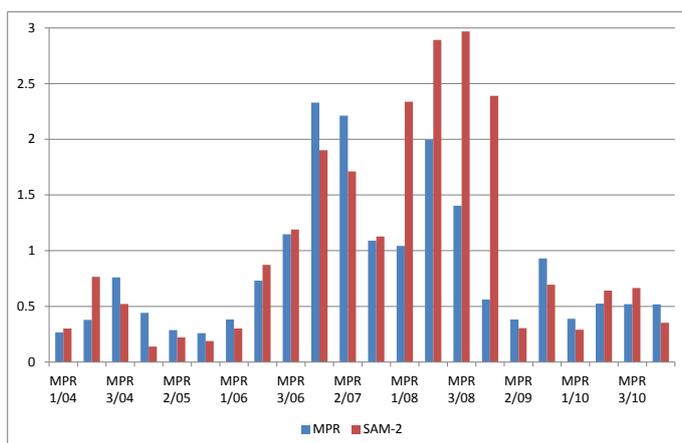


Figure 6. *RMSEs per report, horizons 1 to 4. Seasonally adjusted Mainland-GDP, four-quarter growth. Forecasts from MPR 1/2004 to MPR 3/2010*

(AR(4)). This model is one of the AR models included in SAM. RMSEs for the whole period are calculated in the last three columns. As expected from inspecting Chart 5, forecast errors from MPR and SAM-2 are very similar in the first period. In the second period, MPR forecasts are more accurate than SAM-2 forecasts. Except for the longest horizon, MPR forecast errors are markedly smaller in the later period than in the early period. The forecasts of the AR(4) model are poor compared to the other forecasts.

RMSEs per report are quite similar for SAM-2 and MPR in the early period, see Chart 6. SAM-2 forecasts are somewhat more accurate in more than half of the reports and markedly poorer in only two reports prior to PPR 2/2008. In all the reports in 2008 and the first report in 2009, MPR forecasts are clearly superior to SAM-2 forecasts. When we get past the pronounced downturn in 2008-2009, RMSEs per report are again quite similar and also relatively small.

SAM-2 performs quite well, in particular in periods with relatively stable growth. In periods with large disturbances, adding judgement is crucial in order to improve forecasts of GDP growth.

### 2.2.2 Inflation forecasts

Turning to inflation, forecasts for four-quarter growth in (CPI-ATE) have been published regularly for a number of years. Several statistical models have been employed in the forecasting process. Chart 7 shows MPR forecasts with SAM-2 forecasts and actual inflation rates. For some of the reports, in particular MPR 3/2005, forecasts clearly miss actual developments. SAM-2 forecasts are almost identical to the MPR forecasts for this report. When projections for 2006 were evaluated in MPR 1/2007, Norges Bank pointed to unexpected developments on the supply-side of the economy. High growth in productivity and inward labour migration, as well as strong competition in some product markets, all contributed to decreasing inflation rates.

We also see examples in the early period where forecasts closely follow historical inflation rates, and even indicate turning points a few quarters before they occur. See for instance SAM-2 forecasts in MPR 2/2006.

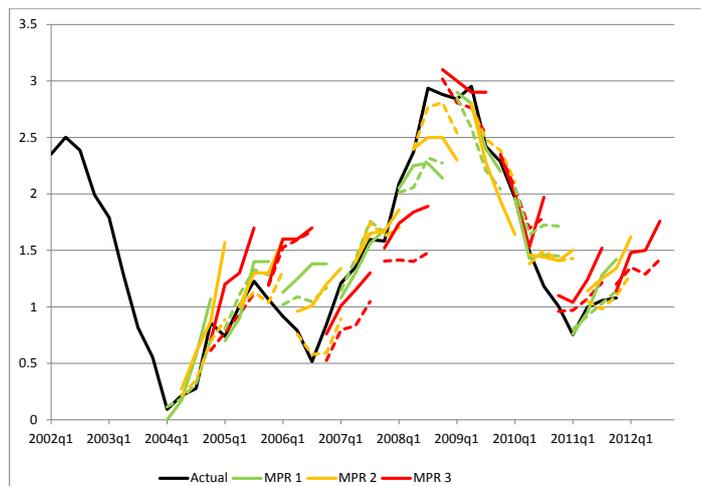


Figure 7. *Seasonally adjusted CPI-ATE. Four-quarter growth rates, per cent. Forecasts from MPR (solid lines) and SAM-2 (dashed lines). Green, yellow and red lines represent MPRs from late March (MPR 1), late June (MPR 2) and late October (MPR 3), respectively. First sequence of forecasts from MPR 1/2004.*

Table 4. RMSE. Forecasts for CPI-ATE, four-quarter growth. Forecasts from MPR 1/2004 to MPR 3/2011. Forecasts in IR/MPR, forecasts using SAM-2 and forecasts with using a monthly AR(12) model. Figures in bold indicate the best-performing forecast.

	MPR 1/04 to MPR 1/08			MPR 2/08 to MPR 3/11			MPR 1/04 to MPR 3/11		
	MPR	SAM-2	AR(12)	MPR	SAM-2	AR(12)	MPR	SAM-2	AR(12)
<b>h=1</b>	<b>0.11</b>	0.14	0.12	0.10	<b>0.09</b>	0.13	<b>0.11</b>	0.12	0.13
<b>h=2</b>	0.33	0.31	<b>0.27</b>	0.21	0.19	<b>0.17</b>	0.29	0.26	<b>0.23</b>
<b>h=3</b>	0.44	0.45	<b>0.41</b>	0.26	<b>0.25</b>	0.31	0.38	0.38	<b>0.37</b>
<b>h=4</b>	0.62	0.61	<b>0.59</b>	0.51	<b>0.41</b>	0.50	0.58	<b>0.54</b>	0.58

In contrast to the case with GDP, RMSEs for SAM-2 forecasts are in general smaller in the period from MPR 2/2008 to MPR 3/2011 than in the earlier period, see Table 4. In the early period, SAM-2 forecasts are slightly more accurate than MPR forecasts for some horizons. In the later period, forecasts from SAM-2 are at least as accurate as the MPR forecasts for all horizons. In particular, SAM-2 was more accurate than MPR for the longest forecast horizon. Forecast accuracy from MPR 2/08 to MPR 3/11 has improved with the new version of SAM, see table 2 in section 2.1.2.<sup>6</sup>

As a benchmark for the performance of inflation forecasts, we have added RMSEs from one of the autoregressive monthly models (AR(12)) that are included in SAM. In the early period, the AR model is slightly more accurate than SAM-2 for all forecast horizons. In the later period, SAM-2 forecasts perform better than the AR model for most of the horizons. Over the period as a whole, the accuracy of the three sets of forecasts are not very different. For 1- and 4-steps ahead forecasts, SAM-2 is slightly more accurate than the AR model, while the AR model is more accurate than SAM-2 for 2- and 3-steps ahead forecasts. Note, however, that the differences are very small

<sup>6</sup>The improvement is understated, since the new version of SAM has been used since MPR 1/2011. Hence, SAM-1 and SAM-2 are identical in the last three reports.

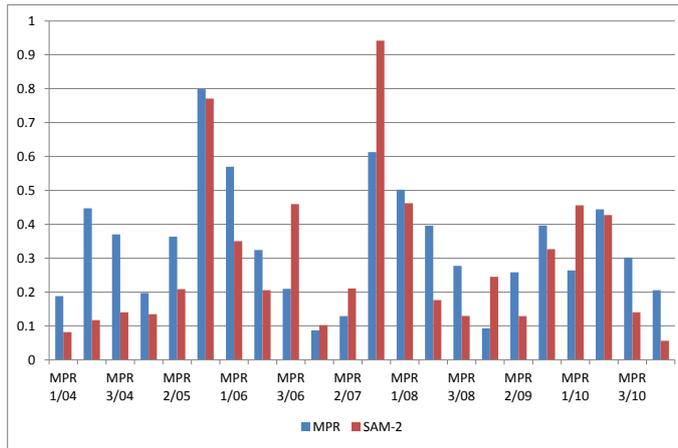


Figure 8. *RMSEs per report, horizons 1 to 4. CPI-ATE, year-over-year growth. Forecasts from MPR 1/2004 to MPR 1/2011*

and probably insignificant. The relatively good performance of the AR model compared with SAM-2 can be explained partly by the fact that the AR model is monthly. In the first 2 issues of MPR every year, we have inflation figures for two of the three months in the first quarter we want to forecast. For a monthly model, that means that two of the three months are known. A quarterly model, on the other hand, forecasts the whole quarter. Hence, monthly models have an information advantage in two of the three reports.<sup>7</sup> The reason SAM-2 does not quite perform equally well, i.e. why is not the best model given (almost) all the weight, can be ascribed to the choice of weighting scheme for the combination of inflation forecasts. Even if one model is superior to all other models, the system attaches weights to other models as well. Other models may be given weights that are large enough for the combined forecast to differ from the forecast from the superior model. This is true for most weighting schemes, to a larger or smaller extent. In SAM, this is more important for inflation forecasts than for GDP growth forecasts. The choice of weights are explained further in [Aastveit et al. \(2011\)](#).

For more details, we again study RMSEs per report, see Chart 8. In most of the

<sup>7</sup>If we only consider the number 3 reports, the performance of SAM-2 is superior to the AR(12) model.

reports, SAM-2 forecasts have been more accurate than the MPR forecasts, particularly at the start of the period but also for the last two reports. Both SAM-2 and MPR display very large forecast errors in a few reports.

The most accurate SAM-2 forecasts were made in connection with MPR 1/2004, MPR 1/2007 and MPR 1/2011. SAM-2 forecasts were also remarkably accurate in MPR 3/2008, anticipating the downward trend in inflation from the beginning of 2009. The worst performing forecasts are made in connection with MPR 3/2005 and MPR 3/2007. We already briefly discussed the first example, where MPR and SAM-2 forecasts both completely missed actual developments. In the latter example, SAM-2 in particular missed the steep increase in inflation rates. Somewhat higher inflation were expected in the *Monetary Policy Report* while SAM-2 predicted the inflation to be stable (see also Chart 7).

### 2.3 Robustness of GDP forecasts

When we evaluate the models and model classes in SAM-2 to compute the weights, we have chosen to use the fifth release of GDP, published one year after the first release, as the actual value that we want to forecast. These growth rates are revised four times and hence should be closer to the “true” value than the first releases.<sup>8</sup> When we evaluate forecasts from SAM-2 and MPR here, we have chosen to evaluate against the first releases of the data. The main reason for choosing the first releases is the limited number of data. Moreover, we interpret the first releases of GDP as rational forecasts of later revisions, justifying taking the first release as our actual value.

Another choice we made in developing SAM-2 was to mainly construct models based on quarter-to-quarter growth rates for GDP. In earlier versions of SAM, forecasts were based on four-quarter growth rates. After forecasting quarter-to-quarter growth rates, four-quarter growth rates are calculated, including densities. When we present forecasts in the *Monetary Policy Reports*, we usually present four-quarter growth rates. These are more easily interpreted than the quarter-to-quarter growth rates, which tend to be

---

<sup>8</sup>See Appendix A in [Aastveit et al. \(2011\)](#) for further discussion.

more volatile. Hence, we evaluate four-quarter growth rates here.

To check the robustness of the evaluation, we have extended the evaluation in two directions: First, we evaluate quarter-to-quarter growth rates. Second, we evaluate both types of growth rates against three different releases, adding the second and the fifth releases to the first release. Finally, we compare the performances of “SAM-2” with the “real-time” SAM-1.

Looking at Table 5, where the upper part is identical to Table 3 (excluding forecasts from the AR model), we observe that results for four-quarter growth rates do not change much with the different vintages, see the middle and bottom parts of the table. In both periods relative forecast performance are quite similar, regardless of which release we evaluate against. In general, forecast errors increase when we evaluate against more revised data.

Turning to quarter-to-quarter growth rates in Table 6, we find some slight differences. In the early period, forecast performances of MPR and SAM are still fairly similar. But MPR forecasts are no longer so consistently performing better in the second than in the first period, regardless of the choice of releases to evaluate against. MPR forecasts still outperforms SAM forecasts, however, but not quite as markedly as with the four-quarter growth evaluation.

In table 7 we have repeated results from the second period for SAM-2, for four-quarter growth and quarter-to-quarter growth, and added RMSEs calculated with SAM-1. Since the weights in SAM-2 are computed on the basis of quarter-to-quarter growth, using fifth releases as actual outcomes, we would expect improvement in forecast accuracy with this combination of growth rates and chosen release of GDP. This is indeed the case, as shown in the bottom right part of the table. Forecast accuracy has also improved for the other alternative releases, but the improvement seems more pronounced when evaluating against fifth release growth rates. Forecast performance is roughly unchanged from SAM-1 to SAM-2 for four-quarter growth rates.

Table 5. RMSE. Forecasts for GDP, four-quarter growth. Horizons 1 to 4. Evaluated against first, second and fifth releases. Forecasts from MPR 1/2004 to MPR 3/2011. Numbers in bold indicate that SAM-2 outperforms MPR.

	MPR 1/04 to MPR 1/08		MPR 2/08 to MPR 2/11		MPR 1/04 to MPR 2/11	
	MPR	SAM-2	MPR	SAM-2	MPR	SAM-2
<b>h=1 1st release</b>	0.55	0.59	0.42	0.78	0.50	0.68
<b>h=2 1st release</b>	1.02	<b>0.99</b>	0.49	1.12	0.82	1.05
<b>h=3 1st release</b>	1.27	1.28	1.01	1.74	1.17	1.50
<b>h=4 1st release</b>	1.37	1.46	1.45	2.29	1.41	1.85
<b>h=1 2nd release</b>	0.79	0.83	0.61	0.95	0.72	0.89
<b>h=2 2nd release</b>	1.16	1.17	0.55	1.23	0.95	1.19
<b>h=3 2nd release</b>	1.36	<b>1.34</b>	1.21	2.01	1.30	1.64
<b>h=4 2nd release</b>	1.47	1.59	1.65	2.53	1.54	2.00
<b>h=1 5th release</b>	0.86	0.87	0.47	0.93	0.74	0.89
<b>h=2 5th release</b>	1.23	<b>1.14</b>	0.74	1.60	1.08	1.32
<b>h=3 5th release</b>	1.31	1.33	1.41	2.38	1.35	1.77
<b>h=4 5th release</b>	1.56	1.75	1.86	2.85	1.66	2.16

Table 6. RMSE. Forecasts for GDP, quarter-to-quarter growth. Horizons 1 to 4. Evaluated against first, second and fifth releases. Forecasts from MPR 1/2004 to MPR 3/2011. Numbers in bold indicate that SAM-2 outperforms MPR.

	MPR 1/04 to MPR 1/08		MPR 2/08 to MPR 3/11		MPR 1/04 to MPR 3/11	
	MPR	SAM-2	MPR	SAM-2	MPR	SAM-2
<b>h=1 1st release</b>	0.39	0.40	0.32	0.41	0.35	0.41
<b>h=2 1st release</b>	0.43	<b>0.38</b>	0.34	0.37	0.40	0.37
<b>h=3 1st release</b>	0.45	0.47	0.45	0.65	0.45	0.55
<b>h=4 1st release</b>	0.37	0.43	0.64	0.73	0.50	0.57
<b>h=1 2nd release</b>	0.51	0.53	0.34	0.52	0.44	0.52
<b>h=2 2nd release</b>	0.48	0.48	0.46	0.54	0.47	0.50
<b>h=3 2nd release</b>	0.52	0.54	0.63	0.87	0.57	0.69
<b>h=4 2nd release</b>	0.56	0.66	0.78	0.86	0.65	0.74
<b>h=1 5th release</b>	0.50	<b>0.45</b>	0.29	0.44	0.43	0.45
<b>h=2 5th release</b>	0.50	<b>0.48</b>	0.60	0.89	0.53	0.65
<b>h=3 5th release</b>	0.59	0.63	0.74	1.01	0.65	0.79
<b>h=4 5th release</b>	0.63	0.73	0.66	0.89	0.64	0.79

Table 7. RMSE. SAM-1 and SAM-2 forecasts for GDP. Evaluated against first, second and fifth releases. Forecasts from MPR 2/2008 to MPR 3/2010. Numbers in bold indicate that SAM-2 outperforms SAM-1.

	Four-quarter growth		Quarter-to-quarter growth	
	SAM-1	SAM-2	SAM-1	SAM-2
<b>h=1 1st release</b>	0.90	0.92	0.52	0.54
<b>h=2 1st release</b>	1.30	1.31	0.56	<b>0.40</b>
<b>h=3 1st release</b>	1.98	2.01	0.83	<b>0.72</b>
<b>h=4 1st release</b>	2.60	<b>2.58</b>	0.94	<b>0.81</b>
<b>h=1 2nd release</b>	1.15	1.18	0.64	0.68
<b>h=2 2nd release</b>	1.44	1.44	0.68	<b>0.57</b>
<b>h=3 2nd release</b>	2.10	2.13	0.96	<b>0.89</b>
<b>h=4 2nd release</b>	2.73	<b>2.71</b>	1.13	<b>0.98</b>
<b>h=1 5th release</b>	1.01	<b>1.01</b>	0.61	<b>0.46</b>
<b>h=2 5th release</b>	1.58	<b>1.57</b>	0.89	<b>0.84</b>
<b>h=3 5th release</b>	2.29	2.35	1.13	<b>1.02</b>
<b>h=4 5th release</b>	3.20	<b>3.16</b>	1.38	<b>1.15</b>

### 3 Concluding remarks

We have evaluated forecasts from SAM and compared them with official forecasts from Norges Bank published in MPR.

In section 2.1 we evaluated actual real-time forecasts from SAM (SAM-1) and MPR from MPR 2/2008 onwards. In this period, SAM became gradually more important in the forecasting process. In section 2.2 we used the new version of SAM (SAM-2) and went back to MPR 1/2004 and computed real-time forecasts for GDP and inflation for the whole period until MPR 3/2011.

Judgement was obviously crucially important in forecasting the downturn in GDP in 2008 and 2009. The dramatic events in the autumn 2008 only showed up later in the data that are input in SAM, and MPR forecasts clearly outperformed SAM forecasts. From mid 2009 forecasts from SAM and MPR were in agreement, with forecast errors of similar magnitude. Inflation forecasts from SAM and MPR were approximately equally accurate. MPR outperformed SAM in 4 of the 9 reports.

Looking at forecasts from early 2004 to early 2008, SAM forecasts perform well compared to MPR forecasts for both GDP growth and inflation. In 9 of the 13 reports from MPR 1/2004 to MPR 1/2008, inflation forecasts from SAM (SAM-2) outperformed MPR forecasts. This result is better than for the later period, where SAM (SAM-1) outperformed MPR forecasts in 5 of 9 reports. This could be an indication that since SAM was incorporated in the MPR forecasting process, inflation forecasts in MPR has improved somewhat.

We have also compared the accuracy of the current SAM (SAM-2) with older versions of the system (SAM-1). The accuracy and hence the usefulness of SAM have increased, in particular for inflation forecasts.

Finally, we have briefly touched upon the issue of revisions of national accounts data, including a robustness check of possible choices of data releases to use as the target, or final outcome. The relative performance of SAM compared with MPR forecasts is largely unchanged for different choices of outcomes.

## References

- Aastveit, K. A., K. R. Gerdrup, and A. S. Jore (2011). Short-term forecasting of GDP and inflation in real-time: Norges Bank's system for averaging models. Staff Memo 2011/9, 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.
- Bjørnland, H. C., K. Gerdrup, A. S. Jore, C. Smith, and L. A. Thorsrud (2012). Does forecast combination improve Norges Bank inflation forecast? *Oxford Bulletin of Economics and Statistics* 74(2), 163–179.
- 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.