EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Kirstin Hubrich, Frauke Skudelny

Forecast combination for euro area inflation: a cure in times of crisis?

No 1972 / October 2016

Abstract

The period of extraordinary volatility in euro area headline inflation starting in 2007 raised the question whether forecast combination methods can be used to hedge against bad forecast performance of single models during such periods and provide more robust forecasts. We investigate this issue for forecasts from a range of short-term forecasting models. Our analysis shows that there is considerable variation of the relative performance of the different models over time. To take that into account we suggest employing performance-based forecast combination methods, in particular one with more weight on the recent forecast performance. We compare such an approach with equal forecast combination that has been found to outperform more sophisticated forecast combination methods in the past, and investigate whether it can improve forecast accuracy over the single best model. The time-varying weights assign weights to the economic interpretations of the forecast stemming from different models. We also include a number of benchmark models in our analysis. The combination methods are evaluated for HICP headline inflation and HICP excluding food and energy. We investigate how forecast accuracy of the combination methods differs between pre-crisis times, the period after the global financial crisis and the full evaluation period including the global financial crisis with its extraordinary volatility in inflation. Overall, we find that forecast combination helps hedge against bad forecast performance and that performance-based weighting outperforms simple averaging.

Keywords: Forecasting, euro area inflation, forecast combinations, forecast evaluation

JEL Classification: C32, C52, C53, E31, E37

Non-technical summary

In this paper we compare the forecast performance of different forecast combination methods with that of a range of short-term inflation forecasting models to evaluate in how far the latter can be used to hedge against bad forecast performance in particular episodes, such as the global financial crisis in 2007/2008. The suite of models that we consider comprises an individual equation framework (an update of Benalal et al. (2004), as also described in ECB (2010)) which is based on individual equations using a number of explanatory variables; a Bayesian Vector Autoregressive (BVAR) model for the five main components of HICP (see Giannone et al. (2014), and also ECB (2010)); Vector Autoregressive (VAR) models of the respective component of the Harmonized Index of Consumer Prices (HICP), with consumer goods PPI, unit labour costs, oil prices, non-oil commodity prices and the nominal effective exchange rate (VAR_C); and a VAR of all 5 HICP components and the aggregate HICP (VAR_U). These models are compared with a range of benchmark models, including an Autoregressive (AR) model, an ARIMA model, a random walk with drift (RWWD) and a random walk as in Atkeson and Ohanian (2001) (RWAO). In addition, we compare them with two univariate models that might capture potential non-linearities, an MA(1) and a STAR model.

We use models that provide forecasts of the five main components of euro area HICP on a monthly basis for the period 1990(1) to 2014(6) and most are updated at least on a quarterly basis. We analyse the results for the full forecast period, but also for the pre-crisis and the post-crisis period and present the evaluation of the aggregation of the component forecasts. We carry out our forecast combination and evaluation based on every third month, as in the quarterly projections of the Eurosystem. We aggregate the disaggregate forecasts to headline HICP inflation forecasts by using the weights that would last have been available and known to the forecaster in real time.

Three different combination approaches are included in our forecast model comparison: the simple average, using equal weights; a performance-based forecast combination using the root mean squared error for a rolling 2-year window of the most recent past to weight the different forecasts; and a performance-based forecast combination as above, with geometrically (backwards) decaying weights, i.e. recent performance is given more weight than performance longer ago.

We find that the best model for forecasting differs depending on whether the overall HICP or the HICP excluding food and energy is considered, and which period and forecast horizon is studied. Therefore we conclude that performance-based forecast combination helps to hedge against bad forecast performance of some of the models in some situations, even though in the presence of large shocks or crises it does not necessarily improve over the best forecast model since the forecast accuracy of all models might, for example, be biased in the same way. Performance-based forecast combination appears to be useful when the models included in the set of models exhibit very different forecast performance over time.

Forecast combination for the full sample period typically improves forecast accuracy over the autoregressive benchmark model for core and headline inflation, and often improves over single multivariate models. The forecast accuracy gain of combinations is largest for inflation excluding energy and food for the full sample. We also find that performance-based forecast combination improves significantly over the simple average for our application.

Investigating combination weights and their development over time, we find significant changes in the weights. Moreover, there appears to be more pronounced evidence for a structural break in the forecast model performance around the time of the recent global financial crisis, in particular for longer horizons. This time-variation in the weights can be seen as capturing non-linearities in the underlying economic relationships. For instance, the importance of certain variables, such as oil prices or labour-market variables, might change over time. In episodes of more volatile inflation, a multivariate model allowing for feedback effects between inflation and its predictors might improve forecast accuracy. The time-variation in the weights assigned to the forecasts from different single models can help to interpret the combination forecast and improve the forecasting models.

Overall, we conclude from our evidence taking into account RMSFE, forecast accuracy tests and turning point predictions that for euro area inflation, first, performance-based combination tends to outperform simple averages, and, second, that performance-based forecast combination protects against bad forecasts from single models, thereby making the forecast more robust.

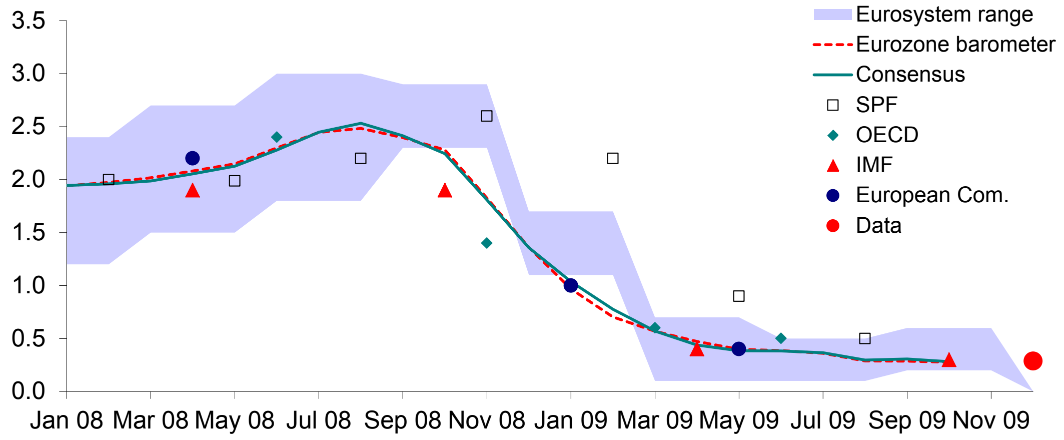
1. Introduction

The period of high volatility in euro area headline inflation starting from around 2007 made it extremely difficult to forecast inflation, even in the short term. Chart 1 shows the evolution of forecasts for annual inflation in 2009 from different institutions and private forecasters. This was a year where it was particularly difficult to forecast inflation, since the volatility of headline inflation was particularly high. The first forecasts represented in the chart have been made in January 2008 and were subsequently revised over time at different time intervals (mostly monthly or quarterly). The last forecast for annual inflation in 2009 was published in October 2009. While the outcome of annual inflation in 2009 was 0.3% (the red dot end-2009), it took until mid-2009 for forecasts of different institutions and private forecasters to come close to this number. Also, the different forecasts appear quite diverse. This raises the question whether forecast combination methods can be used to hedge against bad forecast performance of single models during such periods and provide more robust forecasts.

The increasing number of different models used for short-term inflation forecasting pose a challenge on how to extract and summarise the most important information from the different forecasts in real time. We investigate whether forecast combination methods can help summarise the forecasts from many different models in a meaningful and accurate point forecast. Our analysis shows that there is considerable variation of the relative performance of the different models over time, and this variation can be utilised in a performance-based forecast combination method, in particular one with more weight on the recent forecast performance. Forecast combination with time-varying weights might also be viewed as an approximation of underlying non-linearities. For example, in an environment of relatively stable inflation, a simple autoregressive model might work very well, while in episodes of more volatile inflation, a multivariate model allowing for feedback effects or a model including conditioning information may improve forecast accuracy for inflation.

We investigate how a forecast combination approach with time-varying performance-based combination weights compares with equal forecast combination, which has been found to outperform more sophisticated forecast combination methods in the past, and whether it can improve forecast accuracy over the single best model. We employ a range of single equation and vector autoregressive models built to forecast HICP components and HICP headline inflation and also include a number of benchmark models in our analysis. The combination methods are evaluated for the Harmonized Index of Comsumer Price (HICP) headline inflation and inflation in HICP excluding food and energy in terms of forecast accuracy to investigate the source of good forecast performance and forecast failures in more detail. We investigate to what extent the forecast accuracy of the combination methods differs between normal times before the global financial crisis and the post-crisis period as well as over the full sample including the global financial crisis with its extraordinary volatility in inflation.

Chart 1 Forecasts for annual inflation in 2009



Note: The chart shows annual inflation forecasts (y-axis) forecasted at different points in time (x-axis) by a number of institutions (see legend). The published ranges are shown for the Eurosystem Macroeconomic Projections.

In addition, we provide further evidence on the comparison of a direct forecast of all inflation items in comparison with forecasting the disaggregate components of inflation and aggregating those forecasts and new insights into the role of combinations in that context. This evidence is presented in the Appendix.

The paper is organised as follows: Section 2 presents some theoretical and empirical considerations to motivate forecast combination and discusses the forecast combination methods employed in this paper. Section 3 presents the data and the models used for forecasting inflation as well as the methods of evaluation. Section 4 presents the results of the forecast evaluation for the individual models and for the forecast combination methods. The time-varying performance based forecast combination weights are discussed. Finally, concluding remarks are presented.

2. Forecast combination

2.1. Motivation for forecast combination

The literature on forecast combinationhas recommended combining forecasts in the following situations to get more accurate forecasts:

- models are misspecified in some dimension: forecast combination is more robust against misspecification bias
- information sets underlying the forecasts are sufficiently different and combining information sets might not provide the best forecast
- forecasts from different models will be differently affected by structural breaks

Combining has often been argued to be a useful hedging strategy against structural breaks. Hendry and Clements (2002) argue that this might be the case in the presence of an (unknown) break after the estimation period, in particular when a shift occurs in the intercept of a single variable or a shift of two correlated predictor variables occurs in the opposite direction. Stock and Watson (2004) show empirically that combined forecasts tend to outperform individual forecasts and that simple averages are best in terms of point forecast accuracy. Also, Aiolfi and Timmermann (2006) have shown that an equal-weight combination is best in many situations.

In this study, our motivation to use a forecast combination approach is threefold: First, to investigate whether forecast combination can be a useful hedging strategy against bad forecast performance of single models (see also *ECB monetary policy strategy* 2003) in the presence of high inflation volatility and potential structural shifts during the recent global financial crisis; second, to take into account potential non-linearities and time-variation in the performance of single models and investigate whether that improves forecast accuracy over equal-weight combination schemes; and third, to improve over single forecasts, even though that is not always possible and depends on the particular situation. The performance-based combination weights can potentially also be used in real time to inform the forecaster how well the different models have performed recently relative to each other. They also allow evaluating the relative importance of the underlying economic interpretations of the forecast from the different models.

A combination issue separate from the one just discussed (that focuses on the combination of forecasts of the same variable of interest), is whether combining forecasts of different disaggregate variables to forecast the aggregate ("indirect forecast" of the aggregate) is preferable to combining disaggregate information in a model for the aggregate and use this for forecasting ("direct forecast" of the aggregate). We present some new results on forecasting the aggregate versus aggregating component forecasts in the context of the models employed in this study in Appendix IV.

2.2. Forecast combination methods

In this paper, we include three different combination approaches in our forecast model comparison:

- Simple average (equal weights);
- Performance-based forecast combination using the RMSE for a rolling 2-year window of the most recent past to weight the different forecasts;
- Performance-based forecast combination as above, with geometrically (backwards) decaying weights, i.e. recent performance is given more weight than performance longer ago.

Forecast combination weights are calculated based on forecast accuracy measured in terms of the inverse of the mean squared forecast errors (MSFEs) of the respective models relative to the sum of the MSFEs of all models (see also Timmermann, 2006). A key feature of the performance-based forecast weights using a rolling window is that they vary over time for each of the models, so that the resulting combined forecasts capture potential non-linearities and time-variation in the performance of single models.

Optimal least-squares weights, as discussed, for example, in Timmermann (2006) have some appeal from a theoretical point of view due to their optimality properties. In fact, Genre et al (2010) find for the euro area survey of professional forecasters that optimal least squares weighting combination performs well in comparison with other forecast combination methods. However, optimal least squares weights are known to depend on the precision of the estimate of the variance-covariance matrix of the forecast errors, which might be quite imprecise in short forecast evaluation samples, such as we have in our study. On the same grounds, we do not consider combination methods that are based on estimated non-linear schemes either.

One alternative combination method is performance-based weighting, -which we have chosen for our context - where the weights of the individual models are based on past RMSE performance and therefore are not sensitive to estimation problems in small samples.

$$\hat{y}^{c}_{t+h,t} = \sum_{i=1}^{N} \hat{\omega}_{t+h,t,i}\hat{y}_{t+h,t,i} \,, \qquad \hat{\omega}_{t+h,t,i} = \frac{\left(1/MSE^{K}_{t+h,t,i}\right)}{\sum_{j=1}^{N}\left(1/MSE^{K}_{t+h,t,j}\right)}$$

The schemes we are using also include time-varying weights that give more weight to the recent performance of the models. Such combination schemes are sensible when a time-invariant Gaussian distribution for the forecasts and realisation is not a good approximation which is particularly the case during the recent global financial crisis and the euro-area sovereign debt crisis. Our performance-based combination method that gives more weight to recent performance has some resemblance with the Bayesian approach of dynamic model averaging that has recently been suggested by Koop and Korobilis (2010).[4]

3. Data and forecast models

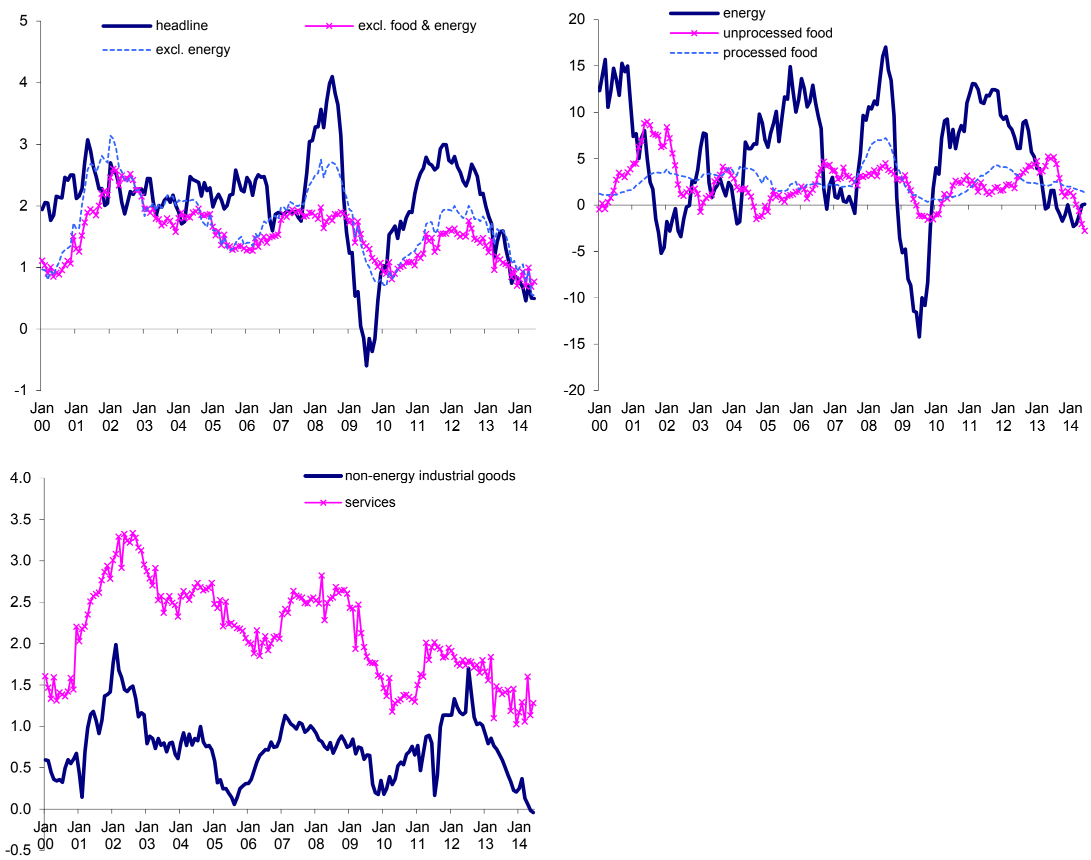
3.1. HICP inflation and its components

The data in this study include headline HICP for the euro area as well as its breakdown into five subcomponents: unprocessed food, processed food, industrial goods, energy and services prices. The data are of monthly frequency, starting in 1990(1) until 2014(6). Chart 2 shows the period of relatively stable headline inflation during most of the 2000s, and the sudden rise in volatility beginning in summer 2007.

These data correspond to what are used in the regular analysis of HICP developments for the euro area as provided, for example, in the commentary of the ECB Economic Bulletin. The breakdown helps the understanding of recent developments and also helps forecasting aggregate HICP developments. It has indeed been shown that in the short run, disaggregated forecasts tend to be more accurate than direct forecasts of the aggregate (see Hubrich (2005), Benalal et al. (2004)).

[4] Koop and Koribilis (2010) refer to Raftery, Karny and Ettler (2010) and Smith and Miller (1986) who argue that to include a “forgetting factor” that imposes an exponential decay for past observations is a sensible and not too restrictive approximation in the context of dynamic model averaging.

Chart 2 HICP headline inflation and components, HICP excluding food and energy



3.2. Forecasting models for inflation

We use models that provide forecasts of the five main components of euro area HICP on a monthly basis and most are updated at least on a quarterly basis. We present the evaluation of the aggregation of the component forecasts.[5] The suite of models that we consider (see Appendix II for a detailed description) comprises an individual equation framework (an update of Benalal et al. (2004), as also described in ECB (2010)) which is based on individual equations using a number of explanatory variables; a Bayesian Vector Autoregressive (BVAR) model for the five main components of HICP (see Giannone et al. (2014), and also ECB (2010)); and VAR models of the respective component of the HICP, with consumer goods PPI, unit labour costs, oil prices, non-oil commodity prices and the nominal effective exchange rate (VAR_C). Forecasts using these models are based on a real time database (see Appendix II for details): they condition on assumptions as produced in the Eurosystem projections.[6] The conditioned VAR (VAR_C) complements the model set since it takes into account only the real side of the

[5] We comment also on the direct forecast of aggregate inflation in comparison with the aggregated component forecast in Appendix IV.

[6] We take the mid-point of the published ranges as the relevant point forecast.

economy and some other important predictors of inflation – like the BVAR – but not the components of HICP inflation.

In addition, we estimate a VAR of all 5 HICP components and the aggregate HICP (VAR_U). This model does not condition on the assumptions in the projections since it does not include any of those variables. We include this model in our forecast combination scheme because it is a relatively simple multivariate model that takes into account the interdependences and second-round effects between the different components of inflation, even though it does not include the real side of the economy as does the BVAR.

We compare these models with two simple univariate non-linear models that might capture potential non-linearities, an MA(1) and a STAR model. The constant parameter MA(1) model from a rolling estimation window is included as an approximation for an Unobserved Component Model with stochastic volatility, as suggested by Stock and Watson (2007). The Smooth Transition Autoregressive (STAR) Model is a generalisation of the 2-regime threshold autoregressive model and allows the estimation of the – potentially smooth - transition from one regime to the other. This model is well-suited for forecasting variables with asymmetric behaviour (see e.g. Teräsvirta, 2006)[7][8]. Pre-crisis studies have presented mixed results on the forecast performance of non-linear models. However, so far there is little evidence on how these models fare during the crisis. We also consider a number of benchmark models in our study, including an AR, an ARIMA model, a random walk with drift (RWWD) and a random walk as in Atkenson and Ohanian (2001) (RWAO).

The lag length for the AR and the VARs is based on the Bayesian information criterion.[9]. The estimation period for those models is determined by the availability of the component data. The first estimation sample is 1992(1) to 2000(1) with the first one-step ahead forecast for 2000(2). The sample is extended recursively until 2014(06). The models are re-estimated for each sample of the recursively expanding estimation window. The forecasts for the BVAR, the VAR, and the individual equation framework are set up so as to reflect the data situation prevailing at the respective, and take into account real-time data for real variables and labour market variables, as well as for expectations regarding exchange rates (flat profile) and oil prices (futures). In that sense, all the models included in the forecast combination are based on the same information set.

We carry out our forecast combination and evaluation based on every third month, as in the quarterly projections of the Eurosystem. We aggregate the disaggregate forecasts to headline HICP inflation forecasts by using the weights that would last have been available and known to the forecaster in real time.

[7] Note that the STAR model is re-estimated only once a year while the other models are re-estimated each month.

[8] Multivariate nonlinear models are beyond the scope of this paper; see for example Hubrich and Teräsvirta (2013) for a recent review of thresholds and smooth transitions in vector autoregressive models. Forecast combinations with time-varying weights might be viewed as an alternative to modelling nonlinearities with multivariate nonlinear models.

[9] Note that we find that models with the lag selection based on the Bayesian information criterion provided more accurate forecasts than when using the Akaike criterion

It is interesting to evaluate whether the combined forecast outperforms the forecasts of some or all of the single forecasting tools and which combination method does best. In addition, in a policy institution, it will be useful to combine forecasts from different inflation forecasting tools to get a single inflation forecast (e.g. in a monthly inflation note or other briefing). The individual models then might still be important to facilitate the interpretation of the combined inflation forecast.

3.3. Forecast evaluation

The evaluation of the above-mentioned models is done in terms of root mean squared error (RMSE). We also present comparisons in terms of mean absolute values (MAD) to investigate the robustness of our results. The evaluation period for the HICP starts with forecasts produced in February 2002 and uses one forecast per quarter until and including May 2007 for the pre-crisis period, from May 2010 to June 2014 for the post-crisis period, and from February 2002 to June 2014 for the full period.

We perform a number of forecast accuracy tests to evaluate the statistical significance of the differences in forecast accuracy between the models. These tests include the Diebold-Mariano (DM) test (Diebold and Mariano, 1995; see also West, 1996) for non-nested models.[10] We also investigate the ability of the models to predict turning points, where turning points are defined as in Canova (2007). A downward turn in inflation occurs when inflation has been below the value at the peak for two periods before the peak and one period after. An upward peak is defined analogously. In addition to the test of *statistical significance,* we also provide graphical evidence on the *economic significance* of the differences in the forecasts and how the relative performance of the models changes over time.

In particular, we evaluate whether different methods of forecast combination improve significantly over the benchmark model in terms of the root mean squared forecast error (RMSE) and help to forecast turning points. We also investigate whether any of the combination methods provides significantly better forecasts than the other methods.

4. Results

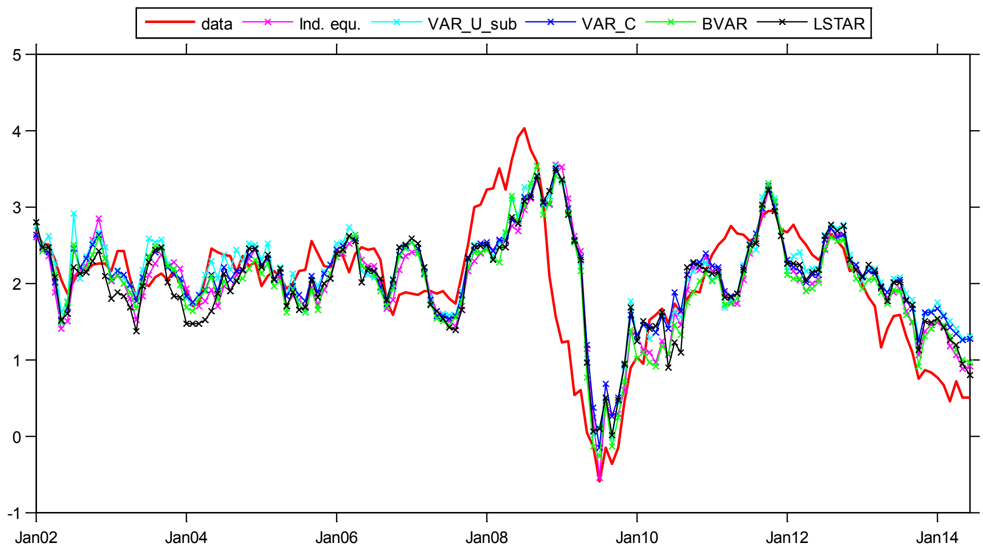
The discussion of the results is structured as follows: We first compare univariate benchmark and non-linear models for HICP headline inflation and HICP inflation excluding food and energy. Second, we compare conditional models and VARs for the same period and variables. In a third step, we present the results on the relative forecast accuracy of the different forecast combination methods for the three above-mentioned periods. For the three different combination

[10] Note that the DM test is applied here since the benchmark direct AR forecast model of aggregate inflation is not nested in the alternative aggregated component models. In further analyses not presented here we have compared the forecasts from the direct forecast of aggregate inflation based on single models with the direct AR forecast using the Clark and West (2007) test for pairwise nested model comparisons and the maximum t-test suggested by Hubrich and West (2010) for a small model sets; see also Granziera, Hubrich and Moon (2014) for an overview of multiple forecast comparison tests. The results did not change the overall conclusions of the paper.

methods – that is simple average, rolling two year window, and geometrically declining performance-based combination - we include two different model sets[11]:

- Combination 1: individual equation framework, BVAR, VAR_U, VAR_C, STAR, AR – here we include the conditional models, the non-linear model (STAR) and the benchmark model.
- Combination 2: individual equation framework, BVAR, ARIMA, STAR, AR – here we only include the four best performing models and the benchmark model.

Chart 3 Headline HICP forecasts – 6 months ahead

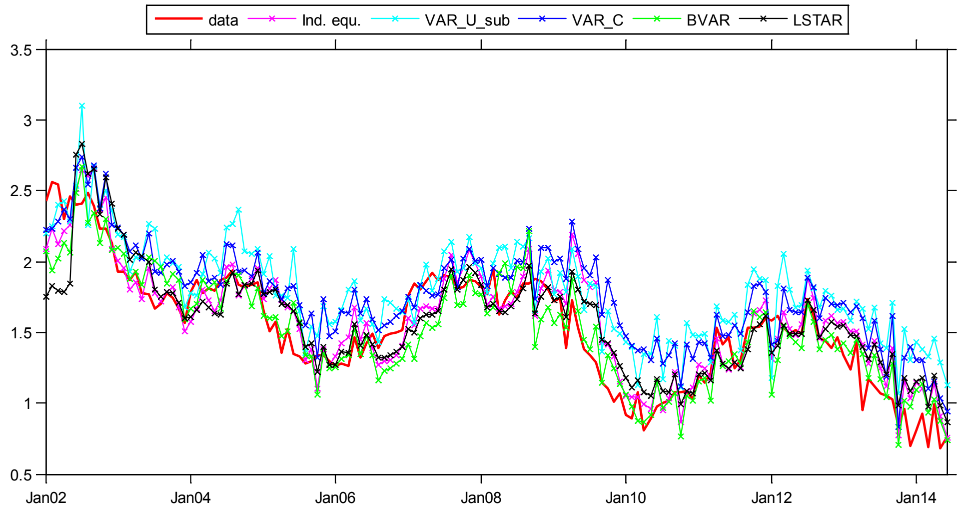


To give a first impression of the forecasts of the different models, Chart 3 shows 6-month ahead forecasts of the individual equation framework, the BVAR, the VARs and of the STAR. All models were lagging in predicting the timing and the size of the increase in headline inflation in 2008 and the decline thereafter. The differences across models are small and typically happen at particular episodes, such as in 2004 or at the end of the sample.

When excluding food and energy, the differences across models appear to be somewhat stronger. Chart 4 shows that the two VARs have overpredicted inflation over most of the sample, while the individual equation framework and the BVAR had no systematic over- or underprediction. This is related to an upward bias in the forecasts for services inflation in the former two models.

[11] A third combination was also investigated, which included the same models as combination 1, but the worst 2 models (corresponding to 20%) were excluded. These excluded models were the RWAO and the MA1.The results from this scheme were very similar than the results for the second combination scheme, and are therefore not presented here.

Chart 4 HICP excluding food and energy forecasts – 6 months ahead



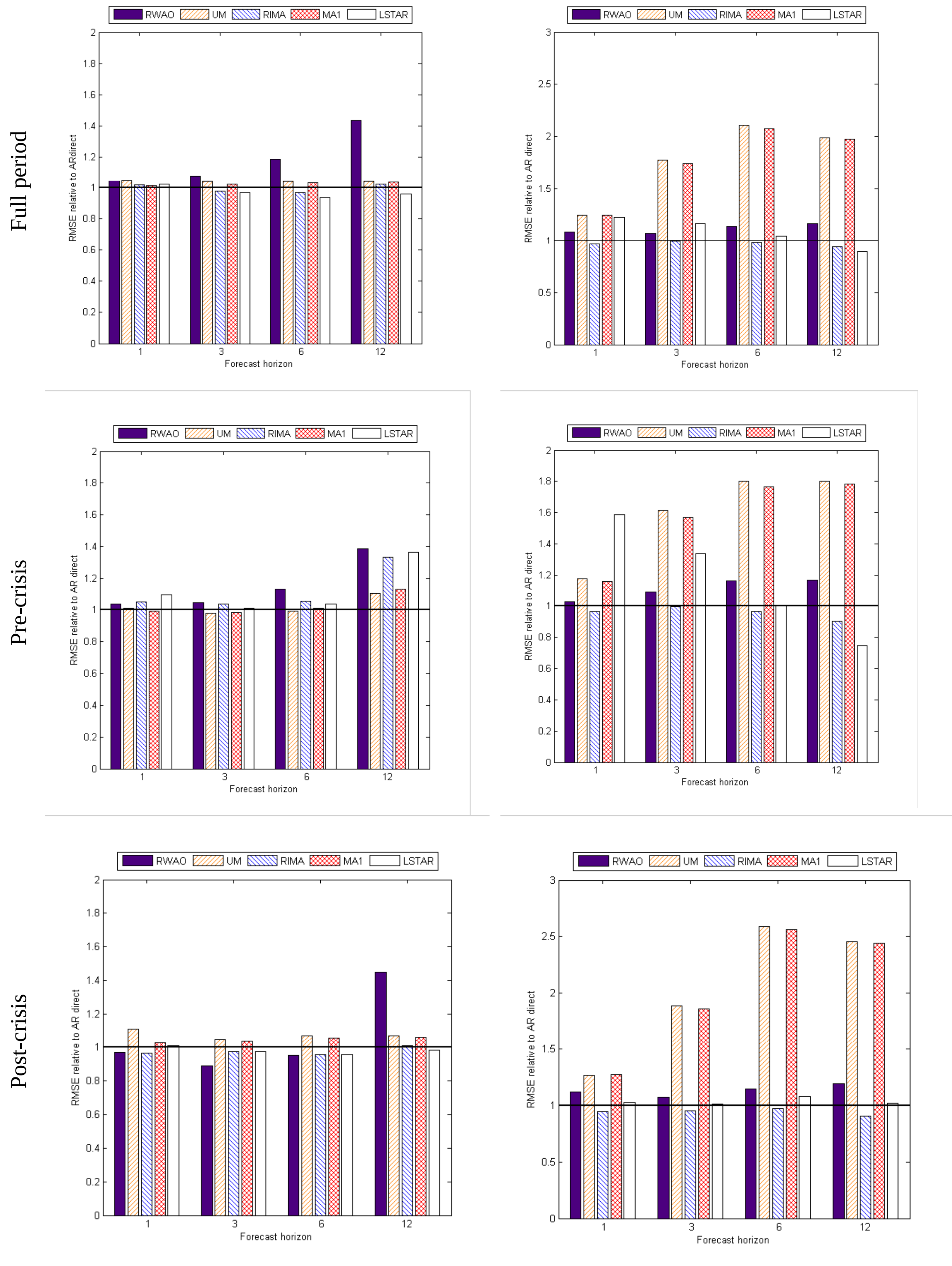
4.1. Univariate benchmark and non-linear model

Chart 5 presents on the left-hand side the RMSEs of univariate and non-linear models relative to the AR benchmark model for total HICP inflation and on the right hand side for HICP inflation excluding food and energy. The first two charts cover the full period, the second two the pre-crisis period and the third two the post crisis period (see Annex I for the full results). A number/bar below one indicates an improvement of the respective model over the AR benchmark.

For total inflation, we find that the RW_AO performs significantly worse than the AR, in particular in the pre-crisis period and over longer horizons also in the post-crisis period. The MA1 based on a rolling estimation window, the random walk with drift (UM) and the ARIMA model also do not systematically outperform the AR. The STAR model shows a somewhat better forecast performance than the AR over the full sample mainly on account of the post-crisis period.

Chart 5 Univariate benchmark and non-linear models

Headline inflation Excl. energy and food



Note: RWAO – random walk as in Atkinson and Ohanian (2001), UM – random walk with drift, RIMA – ARIMA model, MA1 – moving average model of order 1, LSTAR – logistic smooth transition autoregressive model (STAR)

We have investigated how the slope coefficient of the logistic function in the STAR model changes over time. We find that there is a substantial increase in the slope coefficient in 2008 for both headline inflation and inflation excluding energy and food, indicating a rather abrupt change in regime. This might explain the somewhat higher forecast accuracy of the STAR model in comparison with the univariate benchmark models in the post-crisis period.[12] Finally, the ARIMA has a very similar forecast performance as the AR.

Overall, this confirms that the AR model is a useful benchmark model for forecasting headline inflation.

For HICP excluding food and energy, the results are quite different (see Chart 5 right-hand side). While the UM and the MA1 have forecast errors which are sometimes twice as large as the ones of the benchmark, the ARIMA performs better than the benchmark over all horizons. The RWAO model cannot beat the benchmark. The STAR model performs similar or better over longer horizons and for all horizons in the post-crisis period. When considering the pre-crisis period, the STAR performs worse than the benchmark up to 3 steps ahead and is better 12 steps ahead.

We draw two conclusions from this first part of the analysis:

- First, we take the AR model as a benchmark model for the combination forecasts.
- Second, we only include the AR and the STAR model in addition to the individual equations and VAR models in the combination scheme 1, since those were the univariate model with the best performance for headline inflation. Combination scheme 2 also includes the ARIMA, as that performed particularly well for inflation excluding energy and food. We have carried out the forecast comparison also for a combination scheme where we deleted only the 20 % worst models (2 models in our case), but that did not change much the results in comparison to the combination scheme 2.

One reason why the AR benchmark is difficult to beat may be a change in the sum of the coefficients over time, possibly making this model more robust and adaptive to large shifts in inflation such as observed during the crisis. We estimated the AR recursively and found indeed an increase in the sum of the coefficients for headline inflation around the crisis time. However, we could not find the same pattern for inflation excluding food and energy.

4.2. Conditional models and VARs

In this section, we compare different multivariate models. An important difference between these models is that the individual equation framework, the BVAR, and the VAR_C are conditioned on assumptions regarding exchange rates, oil prices, non-oil commodity prices and on interpolated annual variables as used in the Eurosystem projections,[13] while the VAR_U is

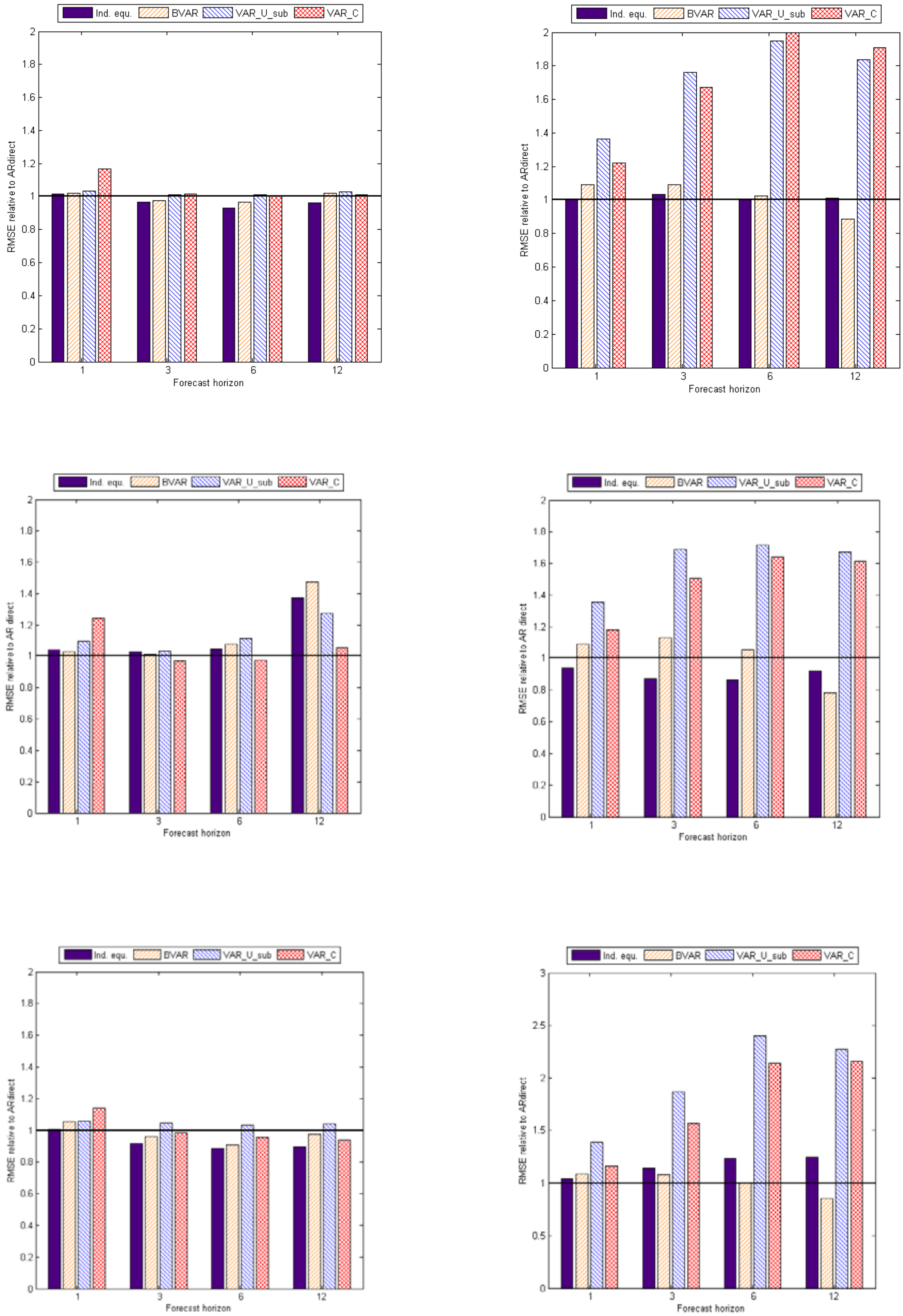
[12] In the past, the record of this kind of model for forecasting has been mixed. So it is interesting that we find some improvement of the STAR forecasts over the benchmark AR forecasts in some situations. These findings indicate that it might be worthwhile to investigate further the potential of non-linear models, in particular multivariate, for forecasting.

[13] We take the mid-point of the published ranges as the relevant point forecast.

Chart 6 Results for conditional models and VARs

Headline inflation | Excl. energy and food

Full period / Pre-crisis / Post-crisis



Note: Ind.equ. – Individual equations, BVAR – Bayesian VAR, VAR_U_sub – unconditional VAR with disaggregate subcomponents of inflation, VAR_C – conditional VAR unconditional as it only includes the HICP and its subcomponents as endogenous variables, i.e. no variable that is part of the assumptions. Therefore, forecast errors of the conditional models are partly a result of errors in the assumptions. However, the assumptions can also help yielding better forecasts.

Chart 6 shows the RMSE relative to the benchmark AR model of the conditional models and the VARs for total HICP inflation (left-hand side) and for HICP inflation excluding food and energy (right-hand side) for the full period, the pre-crisis period and the post-crisis period.

For the full period, most models perform similarly to or worse than the AR benchmark for 1 step ahead forecasts of headline inflation. Beyond that horizon, the individual equations and, though to a lesser extent, the BVAR performed somewhat better. This seems to be mainly related to the post-crisis period where these models show a somewhat better forecast performance. Interestingly, the VAR_C has difficulties in forecasting 1 step ahead while it performs similarly to or better than the AR over the other horizons. All models perform worse than the AR in the pre-crisis period for the 12 step ahead horizon, with particularly large differences for the individual equations, the BVAR and the unconditional VAR_U.

For inflation excluding food and energy (i.e. for the less volatile component of HICP), two of the VAR models, VAR_U and VAR_C, cannot beat the AR in any of the situations, with very large differences compared to the AR in particular beyond the 1 step ahead horizon. The BVAR outperforms the AR model only for the 12-months ahead horizon for all samples. The individual equations have a similar performance as the benchmark over the full period. The subsamples show that this model performs better over all horizons than the AR during the pre-crisis period, while the AR yields lower forecast errors in the post-crisis period. Note that the HICP excluding food and energy consists of the services component and of non-energy industrial goods, which are two components with relatively stable prices. This result shows that the relatively weak forecast performance of the individual equations for headline inflation in the pre-crisis period is mainly related to the difficulty in forecasting energy and food inflation. Meanwhile, the VAR models seem to perform better for the volatile components than for the more stable ones. For the VAR_U, this could be related to the fact that the past movements of all sub-components of the HICP explain the current movement for each of them individually – meaning that core inflation (services and non-industrial goods price inflation) are also explained by food and energy prices which might not be appropriate. For the VAR_C, it could reflect the benefit of using commodity prices as explanatory variable for energy and food prices.

One reason why the individual-equation framework performs relatively better in most situations than the VAR_C (which also includes variables other than the HICP components) is that the individual-equation framework equations for the two components of the HICP excluding food and energy, namely non-energy industrial goods and services, also include taxes. For the latter, no change was assumed when the different forecasts were run. Whether a model forecast is conditional on the Eurosystem projection assumptions seems to matter somewhat as the unconditional VAR_U – the only of the four models which is not conditional - tends to have the highest relative RMSE.

In addition to the RMSE, we have also computed the mean absolute deviation (MAD) to assess the forecast performance of the individual models with a measure that is robust to outliers (see Appendix I third and fourth table). Overall, the results are similar to those obtained with the RMSE. The results are very similar across the univariate models. The errors for the BVAR are a bit higher than for the other models, while those for the individual equations and the conditional VAR are somewhat smaller. Meanwhile, the results for inflation excluding energy and food are somewhat better for the BVAR, while the two VAR models perform less well.

4.3. Combination Results

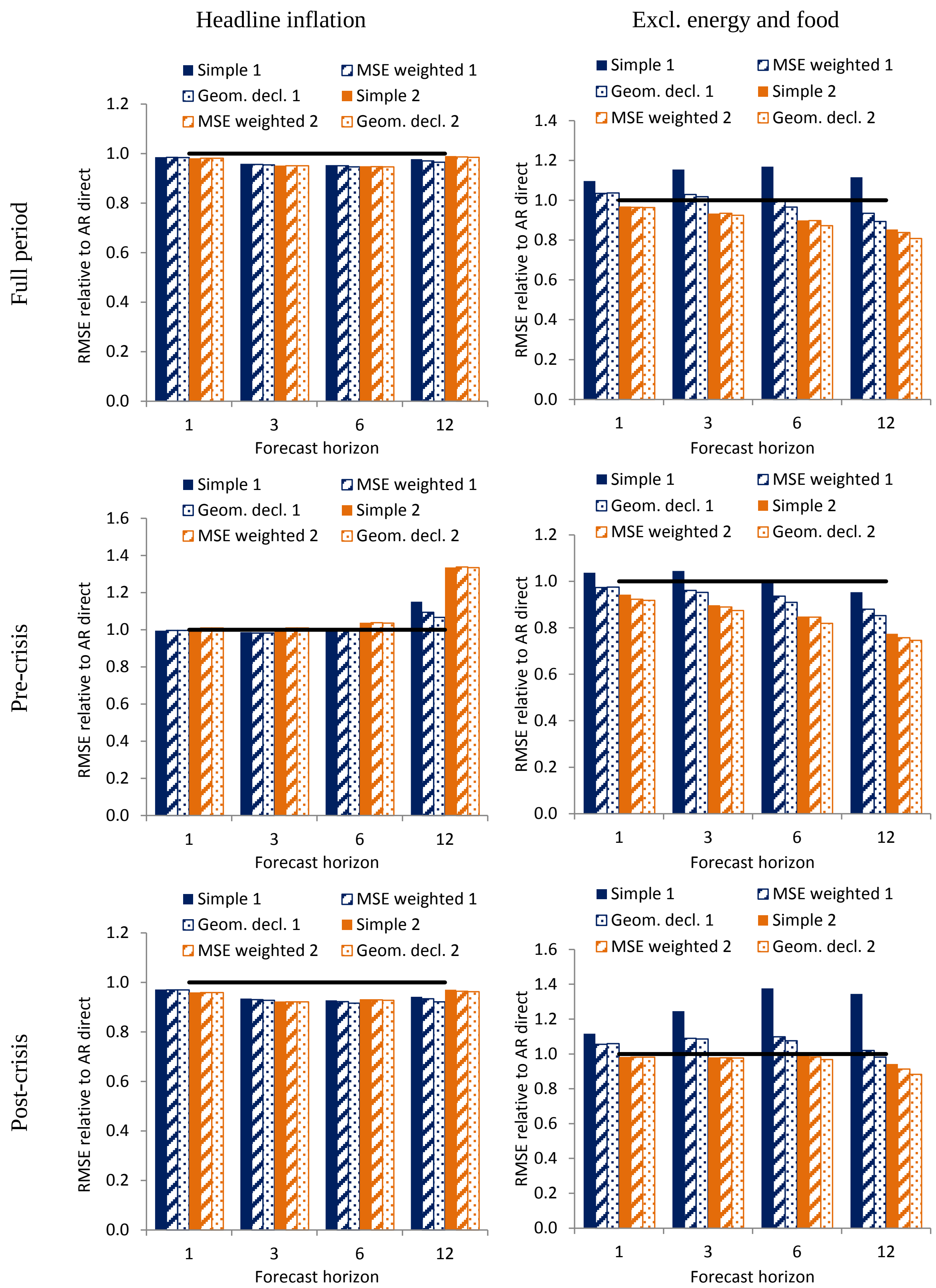
When comparing the different combination schemes for the full period for HICP headline inflation (see first row left-hand side of Chart 7, the combination forecasts perform somewhat better than the AR benchmark model. The differences between the combination methods are rather small. This result is, however, mainly driven by the volatile components, as the forecast performance for the different combinations differs when looking at HICP excluding food and energy inflation (see first row left-hand side of Chart 7). It appears that the second combination scheme, which drops the VAR_C and the VAR_U and adds the ARIMA compared with the first combination scheme, works better than the first combination scheme and also the AR for core inflation. The simple average yields higher forecast errors than the two other weighting schemes, in particular for the first model combination. Taking geometrically declining weights or MSE weights improves the forecast performance over the AR benchmark.

Looking at the pre-crisis period, for headline inflation none of the combination schemes outperforms the AR, while all of them yield lower forecast errors in the post-crisis period. Here again, the differences across models are minor, with the exception of the 12 step ahead forecast where the second combination works worse for the pre-crisis sample. This is consistent with the finding that the VAR models perform relatively better when including the volatile components. The opposite holds true for core inflation, where the combination schemes outperform the benchmark in the pre-crisis period but mostly not in the post-crisis period. The second combination scheme again works somewhat better than the first, with a much stronger difference for the simple average. This is consistent with the finding above that most models outperform the AR in the post-crisis period for headline inflation, but not in the pre-crisis period. It could indicate that the AR is a relatively good model to forecast in particular the volatile component over a period which was mainly characterised by a trend increase in commodity prices, while the individual models which partly include commodity prices work better in a period with higher volatility of commodity prices.

These results are also in line with the mean absolute deviation, i.e. the combination results are similar to or better than the best models, and confirm that forecast combination helps hedge against the worst forecast performance.[14]

[14] We also compared forecasting the aggregate directly and aggregating the component forecasts ('indirect' forecasts) for the two different inflation aggregates we consider in this paper. When looking at the results for the combination methods, we find that forecast accuracy is very similar between the direct and the indirect approach, both for headline and core inflation. We present the results in Appendix IV.

Chart 7 Results forecast combinations



Time-varying Weights in Model combinations

As the RMSEs are average numbers over time, it is useful to provide additional insight by looking at the developments of the performance-based combination weights over time. All charts on the performance based weights for the headline (aggregate) inflation and the HICP excluding food and energy for both combination schemes display considerable variation of the weights over time. We choose to display the weights for combination scheme 2 with backward geometrically declining weights since this scheme improved the forecast accuracy more than the other combination schemes.

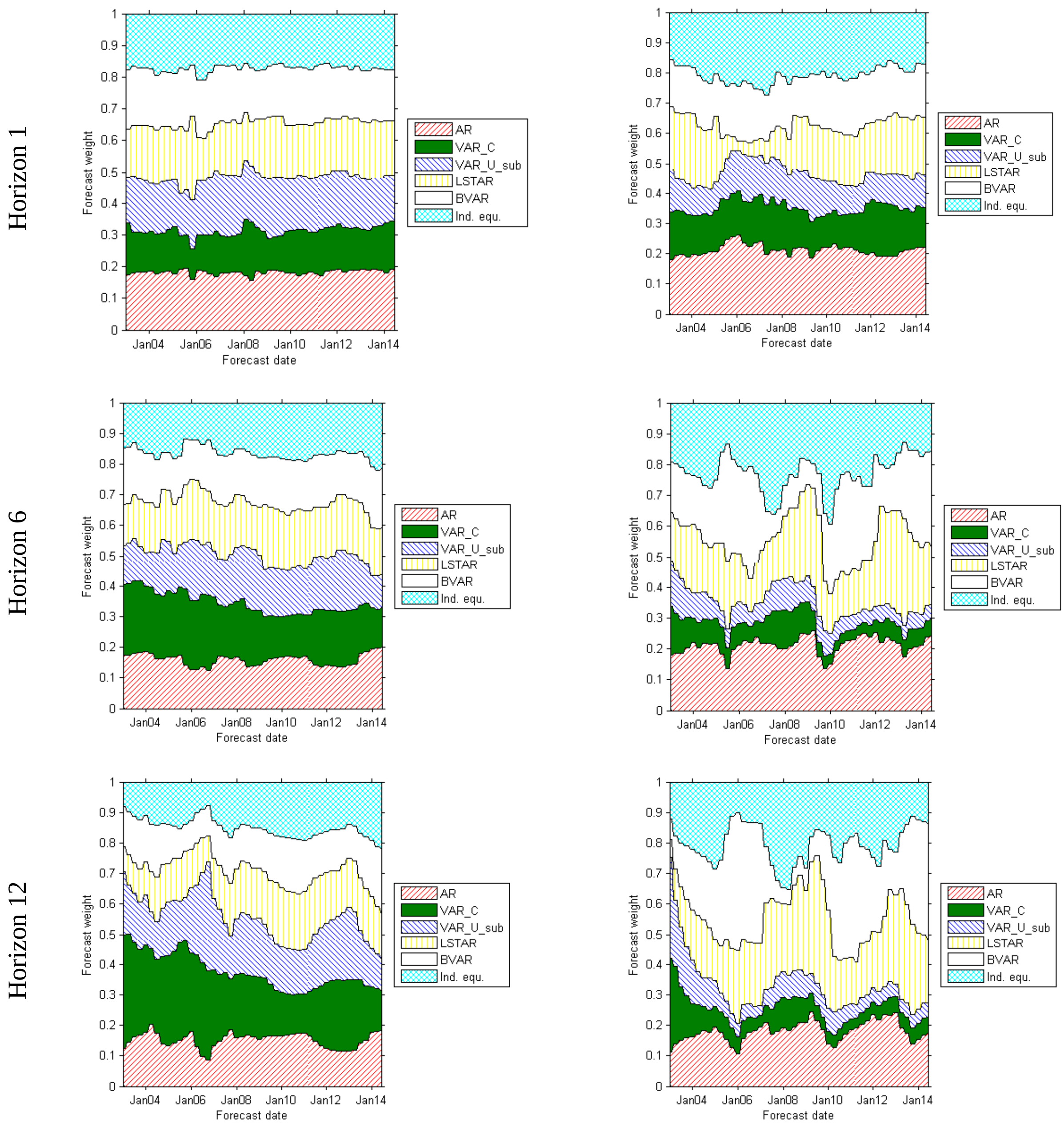
Looking at the performance-based combination weights in Chart 8 shows that there are larger changes in the relative weights starting with the increased volatility in inflation in mid-2007. Even though there are weight changes before, the change in 2007 seems to be somewhat broader across different horizons and models and more pronounced, in particular for the 6- and 12-month horizons. We find that both the rolling estimation window scheme with equal weights within the estimation window (not shown here) and the backward geometrically declining combination weights scheme (within the rolling window) giving recent forecasts more weight than those in the past, exhibit very similar performance. Also, we found similar developments of the relative forecast performance over time represented by the weights for both combination methods. Chart 8 shows that the time-variation in the performance of the models is much more pronounced for inflation excluding energy and food than for headline inflation. This is presumably due to the fact that for headline inflation the forecast performance of all models is negatively affected by the volatile components of inflation, while for inflation excluding energy and food the model choice plays a more important role for the forecast accuracy.

We also carried out Bai-Perron tests for unknown, multiple breakpoints for the performance weights of each of the models in combination scheme 2. The graphs with the significant break dates for the weights of the respective forecasting methods and horizons are presented in Appendix III. The results show a large number of structural breaks in the weights of individual models across all forecast horizons, models and the two aggregate inflation measures considered. The break dates in the weights become more synchronised when the changes are more pronounced, as with longer forecast horizons. This becomes particularly apparent for headline inflation ('hlagg') for a forecast horizon of 12 months where the break dates resulting from the Bai-Perron test are most synchronised across different methods in 2007.

Considering more closely the time-varying performance of the different models for inflation excluding energy and food for a 12-month horizon, we find that in 2007/2008 the BVAR gets substantially less weight than in other periods, while the Individual-Equations and the LSTAR get more weight. This might indicate that the interaction between the different inflation component equations in the BVAR is less important during this time, and that other factor might have played a larger role. This provides an example how the analysis of the time variation in forecast performance helps to interpret the resulting combined forecasts and improve the forecasting models

Chart 8 Performance based weights, Combination 2

Headline inflation   Excl. energy and food



Note: AR – Autoregressive model, VAR_C – conditional VAR, VAR_U_sub – unconditional VAR with disaggregate subcomponents of inflation, LSTAR – logistic smooth transition autoregressive model, BVAR – Bayesian VAR, Ind.equ. – Individual equations

4.4. Forecast Accuracy Tests

To evaluate the statistical significance of the improvement in forecast accuracy of the different models and model combinations we carry out the Diebold-Mariano (DM) test (Diebold and

Mariano, 1996, see also West, 1996) for non-nested forecast model comparisons for headline inflation and inflation excluding energy and food.[15]

In the following, we present the results from the Diebold-Mariano (DM) tests for combination scheme 2, since that scheme performed better in terms of forecast accuracy gains than combination scheme 1.

For the full sample period (see Tables 1a and 1b), the combinations perform best in terms of RMSFE and are often significant, in particular for those with geometrically declining weights. For inflation excluding energy and food this is true for all horizons, while for headline inflation there is only significant improvement for horizons 3 and 6 months.

Table 1a Full sample, headline inflation, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.16		0.81	0.35		0.72	0.56		0.70	0.88		0.70
UM	1.07	-1.18	0.71	1.06	-0.59	0.68	1.06	-0.94	0.66	1.04	-1.22	0.76
MA1	1.03	-0.78	0.77	1.05	-0.21	0.67	1.05	-0.83	0.65	1.04	-1.22	0.78
RIMA	1.03	-1.22	0.78	1.00	0.10	0.72	0.99	-0.16	0.68	1.02	-1.53	0.72
RWAO	1.06	-0.72	0.76	1.10	-0.77	0.71	1.21	-1.30	0.66	1.44	-2.34	0.63
VAR_C	1.18	-2.19	0.72	1.04	0.03	0.76	1.03	0.15	0.71	1.01	-0.28	0.77
VAR_U_sub	1.05	-0.76	0.78	1.03	-0.76	0.68	1.03	-0.88	0.66	1.03	-2.30	0.78
LSTAR	1.04	-0.92	0.77	0.99	0.90	0.70	0.96	0.48	0.64	0.96	-0.47	0.72
BVAR	1.03	-0.81	0.77	1.00	0.99	0.74	0.99	0.66	0.73	1.02	-0.81	0.73
STIP	1.03	-0.46	0.78	0.99	0.47	0.74	0.95	0.38	0.73	0.96	-0.42	0.76
Simple avg	1.00	0.49	0.76	0.98	1.77	0.73	0.97	1.48	0.70	0.98	0.34	0.72
MSE Weighted avg	1.01	0.43	0.76	0.98	1.78	0.73	0.97	1.50	0.70	0.98	0.41	0.73
Geom Decl avg	1.01	0.39	0.77	0.98	1.78	0.73	0.96	1.52	0.70	0.98	0.43	0.73

Note: The different models and their abbreviations are described in Section 3.2. The table presents relative RMSFEs, except for the AR benchmark model where the absolute RMSFE is reported. The t-statistic unadjusted for estimation uncertainty is reported, which is the usual Diebold-Mariano statistic and the critical value for DM test from the t-distribution is 1.282 at a 10% significance level. The accuracy of turning point predictions are evaluated by DC (directon of change or turning point) counting the fraction of times that a change in the right direction was predicted. Combination 2 includes individual equation framework, BVAR, ARIMA, STAR, AR. Green shaded (medium grey) values mark the models with RMSFE smaller than the AR benchmark, orange coloring (light greay) indicates those models that have *significantly* lower RMSFE. Blue shading (dark grey) indicates the best performance in terms of turning points.

[15] Note that the benchmark model is the aggregate AR model, while the alternative model we consider in this table are the aggregated model forecasts of the components. Further results are presented in Appendix IV.

Table 1b Full sample, inflation excl. energy and food, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.08		0.79	0.11		0.76	0.18		0.79	0.37		0.68
UM	1.28	-2.87	0.70	1.77	-4.07	0.57	2.10	-4.29	0.46	1.94	-3.79	0.45
MA1	1.28	-2.68	0.69	1.74	-4.19	0.58	2.07	-4.25	0.46	1.93	-3.73	0.45
RIMA	0.99	1.63	0.80	0.99	-1.13	0.75	0.97	-1.28	0.80	0.92	-1.28	0.78
RWAO	1.11	-0.79	0.77	1.07	-1.09	0.74	1.13	-1.99	0.72	1.14	-2.17	0.39
VAR_C	1.40	-3.13	0.70	1.75	-4.28	0.59	1.94	-3.97	0.46	1.79	-3.32	0.43
VAR_U_sub	1.26	-3.65	0.69	1.67	-5.63	0.62	1.99	-4.70	0.53	1.87	-3.30	0.45
LSTAR	1.26	-1.21	0.75	1.16	-1.58	0.70	1.04	-1.08	0.75	0.87	-0.02	0.76
BVAR	1.13	-1.71	0.74	1.09	-1.68	0.73	1.02	-0.53	0.75	0.87	0.79	0.76
Ind. equ.	1.03	0.91	0.82	1.03	0.29	0.74	1.00	0.70	0.75	0.99	1.03	0.62
Simple avg 2	1.01	0.98	0.83	0.96	1.08	0.77	0.90	1.67	0.80	0.84	2.66	0.80
MSE Weighted avg 2	1.00	1.24	0.83	0.95	1.18	0.75	0.90	1.74	0.80	0.82	3.02	0.79
Geom Decl avg 2	1.00	1.22	0.81	0.95	1.35	0.75	0.88	2.17	0.81	0.79	3.38	0.82

Note: See Table 1.

Table 2a Pre-crisis sample, headline inflation, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.14		0.79	0.26		0.65	0.32		0.67	0.36		0.61
UM	0.97	0.54	0.71	0.95	1.73	0.68	0.93	1.87	0.65	0.83	0.70	0.72
MA1	0.96	0.83	0.74	0.95	1.71	0.67	0.94	1.52	0.58	0.85	0.45	0.74
RIMA	1.02	-1.52	0.79	1.00	-1.31	0.68	0.99	-1.96	0.60	1.00	-2.32	0.63
RWAO	1.00	-0.82	0.74	1.02	-0.47	0.65	1.06	-1.80	0.52	1.04	-3.38	0.52
VAR_C	1.20	-3.29	0.72	0.94	1.36	0.71	0.91	1.32	0.65	0.79	-0.03	0.69
VAR_U_sub	1.06	-0.79	0.77	1.00	-0.77	0.68	1.05	-1.87	0.58	0.96	-2.64	0.72
LSTAR	1.06	-1.42	0.75	0.98	0.15	0.65	0.97	-1.31	0.58	1.02	-1.87	0.57
BVAR	1.01	-0.68	0.75	1.00	-0.15	0.68	0.98	-1.32	0.62	1.03	-2.42	0.59
STIP	0.99	-0.87	0.79	0.98	-0.84	0.70	1.01	-1.66	0.68	1.11	-2.05	0.61
Simple avg	0.99	-0.80	0.75	0.97	-0.03	0.68	0.96	-0.30	0.65	0.98	-1.32	0.59
MSE Weighted avg	0.99	-0.90	0.75	0.97	-0.06	0.68	0.96	-0.31	0.67	0.98	-1.32	0.59
Geom Decl avg	0.99	-0.94	0.75	0.97	-0.05	0.68	0.96	-0.31	0.67	0.98	-1.32	0.59

Note: See Table 1.

Table 2b Pre-crisis sample, inflation excl. energy and food, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.06		0.80	0.10		0.71	0.17		0.78	0.35		0.72
UM	1.22	-0.74	0.77	1.66	-1.98	0.67	1.86	-2.11	0.52	1.81	-1.36	0.54
MA1	1.21	-0.43	0.74	1.62	-2.01	0.67	1.82	-2.03	0.53	1.79	-1.29	0.56
RIMA	1.00	1.06	0.77	1.03	-1.33	0.70	1.00	-1.45	0.78	0.91	-1.59	0.78
RWAO	1.07	1.10	0.75	1.13	-0.62	0.68	1.20	-1.36	0.72	1.17	-1.46	0.44
VAR_C	1.23	-1.20	0.72	1.55	-2.25	0.68	1.69	-1.95	0.57	1.62	-0.85	0.50
VAR_U_sub	1.41	-1.17	0.72	1.74	-2.90	0.67	1.77	-1.99	0.63	1.68	-0.71	0.56
LSTAR	1.65	-1.13	0.69	1.38	-1.37	0.70	1.03	-0.81	0.72	0.75	-0.16	0.72
BVAR	1.14	-0.34	0.68	1.17	-1.50	0.70	1.09	-0.61	0.70	0.78	0.76	0.78
Ind. equ.	0.98	2.09	0.83	0.90	1.57	0.78	0.89	1.58	0.80	0.92	1.71	0.72
Simple avg 2	1.02	0.53	0.79	0.97	0.78	0.75	0.89	1.35	0.83	0.79	2.03	0.80
MSE Weighted avg 2	1.00	1.11	0.79	0.96	1.03	0.70	0.89	1.41	0.82	0.77	2.15	0.83
Geom Decl avg 2	1.00	1.17	0.77	0.95	1.19	0.70	0.86	1.70	0.83	0.75	2.27	0.82

Note: See Table 1.

Table 3a Post-crisis sample, headline inflation, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.13		0.82	0.32		0.69	0.46		0.67	0.78		0.59
UM	1.16	-1.71	0.70	1.10	-1.75	0.60	1.14	-2.08	0.53	1.07	-1.84	0.69
MA1	1.07	-1.58	0.78	1.09	-1.86	0.58	1.12	-2.19	0.58	1.06	-1.89	0.69
RIMA	1.00	0.79	0.76	1.02	1.70	0.69	1.01	1.54	0.62	1.01	-0.04	0.62
RWAO	1.01	-0.76	0.78	0.93	-0.72	0.65	1.01	-1.11	0.71	1.45	-2.40	0.59
VAR_C	1.18	-1.53	0.74	1.04	-0.79	0.73	1.01	-0.54	0.69	0.94	-0.40	0.74
VAR_U_sub	1.10	-1.27	0.74	1.10	-1.31	0.58	1.10	-0.05	0.56	1.04	-1.60	0.72
LSTAR	1.05	0.35	0.76	1.02	1.34	0.63	1.01	1.38	0.58	0.98	1.21	0.72
BVAR	1.09	-0.96	0.80	1.01	1.29	0.71	0.96	1.93	0.73	0.97	0.45	0.77
STIP	1.05	-0.47	0.74	0.96	1.56	0.67	0.94	2.30	0.69	0.89	1.46	0.80
Simple avg	1.04	0.54	0.76	1.00	1.25	0.69	0.98	1.02	0.67	0.93	0.33	0.69
MSE Weighted avg	1.02	0.53	0.76	0.99	1.28	0.69	0.97	1.05	0.67	0.95	0.38	0.74
Geom Decl avg	1.02	0.53	0.76	0.99	1.28	0.69	0.97	1.07	0.67	0.95	0.40	0.74

Note: See Table 1.

Table 3b Post-crisis sample, inflation excl. energy and food, combination scheme 2 – relative RMSFE, DM test, turning point prediction

Horizon	1			3			6			12		
	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC	RMSFE rel.	t-stat adj	DC
AR (benchmark)	0.09		0.78	0.12		0.83	0.17		0.82	0.35		0.64
UM	1.30	-3.92	0.68	1.85	-4.85	0.56	2.51	-4.98	0.44	2.36	-5.11	0.44
MA1	1.31	-3.52	0.68	1.83	-5.00	0.56	2.48	-5.02	0.44	2.34	-5.12	0.44
RIMA	0.97	0.75	0.82	0.94	-0.11	0.81	0.94	-0.09	0.89	0.87	1.33	0.92
RWAO	1.15	-1.94	0.76	1.05	-2.00	0.79	1.11	-2.35	0.73	1.14	-2.26	0.36
VAR_C	1.19	-3.07	0.72	1.54	-4.26	0.60	2.07	-4.36	0.49	2.07	-4.67	0.44
VAR_U_sub	1.42	-4.03	0.66	1.84	-5.36	0.60	2.33	-5.93	0.47	2.18	-5.57	0.44
LSTAR	1.06	-1.77	0.82	0.99	-1.30	0.73	1.04	-1.01	0.82	0.98	0.34	0.82
BVAR	1.11	-2.04	0.82	1.06	-0.79	0.83	0.97	0.12	0.84	0.82	1.63	0.87
Ind. equ.	1.07	-0.70	0.80	1.12	-1.53	0.71	1.19	-0.77	0.78	1.19	-0.75	0.54
Simple avg 2	1.00	0.89	0.86	0.96	0.57	0.81	0.95	0.29	0.82	0.89	1.13	0.90
MSE Weighted avg 2	1.00	0.90	0.88	0.96	0.60	0.81	0.95	0.37	0.82	0.86	1.66	0.90
Geom Decl avg 2	1.00	0.88	0.86	0.95	0.60	0.81	0.93	0.57	0.82	0.84	2.06	0.90

Note: See Table 1.

For the pre-crisis period, the combinations also significantly outperform the single models for inflation excluding energy and food over the longer horizon (see Table 2b), with the individual equations also significantly improving the forecast performance for most horizons. For headline inflation, the MA(1) model, the random walk with drift (UM) and the conditional VAR significantly outperform the AR model (see Table 2a).

For the post-crisis period, the combinations and the ARIMA and the BVAR model outperform the AR model for inflation excluding energy and food over a 12-month horizon only (see Table 3b), while for headline inflation there is a significant improvement for the combinations and the individual equation framework over a 3-month horizon (see Table 3a). The individual equation framework also significantly improves over the AR model for the 6- and 12-month horizons.

Overall, these results suggest that forecast combination leads to significant improvements in terms of RMSFE for core inflation, and occasionally also for headline inflation except for the pre-crisis period. For headline inflation during the pre-crisis period, simple univariate linear models like the RW with drift and the MA(1) are usually significantly better than the AR model, while for the post-crisis period it is the BVAR and the individual equation framework that improve significantly over the AR model.[16] This result is in line with results from the literature that during the Great moderation (pre-crisis) it was very difficult to improve over simple univariate models since there was not much variation in inflation to be explained (see e.g. Stock and Watson, 2008, and Hendry and Hubrich, 2011). On the other hand, the combinations outperform the AR model over forecast horizons of 3, 6 and 12 months for inflation excluding

[16]These test results are confirmed when carrying out the tests proposed by Hubrich and West (2011), which take into account that a small number of forecast models are compared here instead of only pairwise comparisons of forecasts by using the appropriate larger critical values. The results are available from the authors upon request.

energy and food for the full and the pre-crisis sample. This improvement is significant for 6- and 12-month horizons and for the full sample also for a 3- month horizon. More generally, our evidence suggests using combination schemes, because according to our full sample results, combination forecasts improve significantly over the autoregressive benchmark model for both headline inflation and inflation excluding energy and food. For the pre- and post-crisis episodes, combinations perform at least as well as the benchmark or improve over it, and the improvements are often significant.

We have also carried out DM tests comparing the different combination schemes (see Table 4). We find that the MSE-weighted combination scheme outperforms the simple average for a forecast horizon of 1 month, but for horizons of 6 and 12 months, the geometrically declining weights combination forecasts outperform the simple average. Comparing the MSE-weighted combination scheme with the geometrically declining one (not presented in the tables, but available upon request), we find that the latter combination scheme outperforms the former only for headline inflation and for a 1-month horizon, but otherwise the two forecast combination methods are not significantly different from each other. It is plausible that giving higher weight to models that do best in the most recent period improves the short-term horizon forecast.

Table 4: DM test, different forecast combinations vs simple average

		MSE weighted	Geom. decl. weights
Headline inflation	h=1	1.745	-1.298
Inflation excl. energy and food	h=1	-1.185	-0.589
Headline inflation	h=6	-1.237	2.039
Inflation excl. energy and food	h=6	0.277	2.906
Headline inflation	h=12	-2.524	2.859
Inflation excl. energy and food	h=12	-1.194	3.405

Note: t-statistics reported; forecast horizons are h=1,6 and 12 months, yellow shading indicates significantly better weighting scheme; benchmark model for this comparison is the simple average; combination scheme 2 is underlying this comparison, i.e. the four best forecast models and the AR benchmark are included in the forecast combinations.

4.5. Turning Point Predictions

The accuracy of the turning point predictions of the different models is evaluated by the fraction of times forecasts predicted a change in the right direction, denoted "DC" in the tables. As can be seen in Table 1b, over the full sample, for inflation excluding energy and food, the forecast combinations perform best in terms of turning point prediction, whereas for headline inflation, the VARs (including VAR_C, VAR_U_SUB, BVAR) and the individual equation framework tend to perform best (see Table 1a), where the exact best model depends on the horizon.

For the pre-crisis period (Table 2a and 2b), the combinations also outperform the single models for the longer horizon in terms of turning point prediction for inflation excluding energy and food. Otherwise, it is the individual-equation framework that performs best. For headline inflation, the best models include the individual-equation framework, VAR_C and MA(1).

For the post-crisis period (Table 3b), the combination, BVAR and ARIMA models do best for inflation excluding energy and food in terms of turning point prediction, depending on the horizon, while for headline inflation, the AR, the VAR_C, the BVAR or the individual equation framework provide the most accurate turning-point prediction depending on the horizon.

Overall, forecast combinations outperform single models for inflation excluding energy and food for all horizon over the full sample, while for headline inflation, usually single models prevail, but which particular single model performs best changes with the forecast horizon and forecast evaluation period considered. Since combinations usually still perform better than many of the single models, and there is some variation concerning which model performs best over time and horizons, we conclude from the turning-point prediction evidence that combining forecasts still hedges against bad performance of some models.

5. Conclusions

The extraordinary volatility in inflation during the global financial crisis led to large forecast errors of inflation forecasting models. This raised the question whether forecast combination methods can help to improve forecast performance and/or provide more robust forecast performance in the presence of large shocks.

We provide evidence that forecast combination methods have helped in the crisis period for the models considered here. We find that: 1. performance-based combination methods tend to perform at least as well and, in a number of cases, significantly better than the AR benchmark model; 2. forecast combination methods also outperform a large number of the alternative models; 3. the relative forecast performance is particularly good for headline inflation in the post-crisis period, and for core inflation in the pre-crisis period. For the full sample, which includes substantially different episodes in terms of inflation dynamics, forecast combination methods exhibit comparatively high forecast accuracy for both headline and core inflation; 4. Forecast combination for the full sample period typically improves forecast accuracy over the autoregressive benchmark model for core and headline inflation, and often improves over single multivariate models. The forecast accuracy gain of combinations is largest for inflation excluding energy and food for the full sample; 5. We find that performance-based forecast combination improves significantly over the simple average for our application..

We find that the best model for forecasting differs depending on whether the overall HICP or the HICP excluding food and energy is considered, and which period and forecast horizon is studied. Therefore we conclude that performance-based forecast combination helps to hedge against bad forecast performance of some of the models in some situations, even though in the presence of large shocks or crises it does not necessarily improve over the best forecast model since the forecast accuracy of all models might, for example, be biased in the same way. Performance-based forecast combination appears to be useful when the models included in the set of models exhibit very different forecast performance over time.

Investigating combination weights and their development over time, we find significant changes in the weights. Moreover, there appears to be more pronounced evidence for a structural break in the forecast model performance around the time of the recent global financial crisis, in particular for longer horizons. This time-variation in the weights can be seen as capturing non-linearities in the underlying economic relationships. For instance, the importance of certain variables, such as oil prices or labour-market variables, might change over time. In episodes of more volatile inflation, a multivariate model allowing for feedback effects between inflation and its predictors might improve forecast accuracy. The time-variation in the weights assigned to the forecasts from different single models can help to interpret the combination forecast and improve the forecasting models.

Overall, we conclude from our evidence taking into account RMSFE, forecast accuracy tests and turning point predictions that for euro area inflation, first, performance-based combination tends to outperform simple averages, and, second, that performance-based forecast combination protects against bad forecasts from single models, thereby making the forecast more robust.

References

Aiolfi, M. and Timmermann, A. (2006). Persistence in Forecasting Performance and Conditional Combination Strategies. *Journal of Econometrics*. Vol. 135 (1-2). p 31-53.

Atkeson, A., and Ohanian, L. E. (2001), Are Phillips Curves Useful for Forecasting Inflation? *Federal Reserve Bank of Minneapolis Quarterly Review*, 25 (1), 2–11.

Banbura, M., Giannone, D. and Reichlin, L. (2010), Large Bayesian VARs, *Journal of Applied Econometrics*, vol. 25(1), pages 71-92.

Benalal, Diaz Del Hoyo, Landau, Roma and Skudelny (2004), To aggregate or not to aggregate? Euro area inflation forecasting, ECB WPS, 374.

Canova, F. (2007), G-7 inflation forecasts: Random walk, Phillips curve or what else?, *Macroeconomic Dynamics*, 11, 2007, 1–30.

Clark, T. E. and West, K. D. (2007), Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics* 138(1): 291–311.

Clements, M. P. and Hendry, D. F. (2004), Pooling forecasts, *Econometrics Journal* 7: 1–31.

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253– 263.

ECB (2010), Tools for preparing short-term projections of euro area inflation. Monthly Bulletin, April.

Genre, V., Kenny, G., Meyler, A. and Timmermann, A. (2010), Combining Expert Forecasts: Can Anything Beat the Simple Average?. *International Journal of Forecasting* 2013, 29, 108-121

Giannone. D., Lenza, M. and G.E. Primiceri (2015), Prior Selection for Vector Autoregressions, The Review of Economics and Statistics, vol. 97(2), 436-451

Giannone, D., Lenza, M., Momferatou, D. and Onorante, L. (2014), Short-Term Inflation Projections: a Bayesian Vector Autoregressive approach, International Journal of Forecasting,, vol. 30(3), 635-644.

Hubrich, K. (2005). Forecasting euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?, *International Journal of Forecasting* 21(1): 119–136.

Granziera, E., Hubrich, K. and R. Moon (2014), A predictability test for a small number of nested models, *Journal of Econometrics*, Volume 182, Issue 1, 174–185

Hendry, D. F. and Hubrich, K. (2006). Forecasting aggregates by disaggregates, European Central Bank Working Paper 589.

Hendry, D. F. and Hubrich, K. (2011). Combining disaggregate forecasts or disaggregate information to forecast an aggregate?, Journal of Business and Economics Statistics, 29(2): 216-227.

Hubrich, K. and Teräsvirta, T. (2013) Thresholds and smooth transitions in vector autoregressive models, Advances in Econometrics "VAR Models in Macroeconomics, Financial Econometrics, and Forecasting", Vol. 31, 2013

Hubrich, K. and West, K. (2010). Forecast evaluation of small nested model sets, *Journal of Applied Econometrics, 25(4):. 574-594*

Koop, G. and D. Korobilis (2012), Forecasting Inflation Using Dynamic Model Averaging, International Economic Review, Vol. 53, No. 3, August 2012

Raftery, A., M. Karny, and P. Ettler (2010), Online Prediction under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill, Technometrics 52 (2010), 52–66.

Smith, J., and J. Miller (1986), A Non-Gaussian State Space Model and Application to Prediction Records, Journal of the Royal Statistical Society, Series B 48 (1986), 79–88.

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

Stock, J.H. and Watson, M. (2007), Why Has U.S. Inflation Become Harder to Forecast?, *Journal of Money, Banking and Credit*, Vol. 39, No. 1, February 2007

Teräsvirta, T. (2006), Forecasting economic variables with nonlinear models, in: *Handbook of Economic Forecasting*, eds Elliott, G., Granger C. and A. Timmerman

Timmerman, A. (2006), Forecast Combination, in: *Handbook of Economic Forecasting*, eds Elliott, G., Granger C. and A. Timmerman

West K.D. (1996), Asymptotic inference about predictive ability. Econometrica, Vol. 64, 1067 – 1084

Appendix I Forecast errors (RMSE and MAD)

RMSE - HICP - full period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom decl avg 2
1	0.17	0.17	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.19	0.16	0.16	0.16	0.16	0.16	0.16
2	0.28	0.28	0.26	0.27	0.27	0.26	0.26	0.26	0.27	0.27	0.26	0.26	0.26	0.26	0.26	0.26
3	0.39	0.37	0.35	0.35	0.37	0.35	0.35	0.35	0.36	0.37	0.35	0.35	0.35	0.35	0.35	0.35
4	0.48	0.45	0.42	0.42	0.44	0.41	0.41	0.42	0.44	0.44	0.42	0.41	0.41	0.41	0.41	0.41
5	0.58	0.53	0.49	0.49	0.52	0.47	0.47	0.49	0.51	0.51	0.48	0.48	0.48	0.48	0.48	0.48
6	0.68	0.59	0.56	0.55	0.59	0.53	0.53	0.55	0.57	0.57	0.54	0.54	0.54	0.54	0.54	0.54
7	0.77	0.66	0.62	0.62	0.65	0.59	0.59	0.61	0.64	0.63	0.60	0.60	0.60	0.60	0.60	0.60
8	0.88	0.72	0.68	0.68	0.71	0.65	0.64	0.67	0.70	0.69	0.66	0.66	0.66	0.66	0.66	0.66
9	0.97	0.77	0.74	0.74	0.77	0.71	0.70	0.73	0.75	0.75	0.71	0.71	0.71	0.72	0.72	0.71
10	1.07	0.82	0.79	0.80	0.82	0.75	0.75	0.78	0.80	0.80	0.76	0.76	0.75	0.77	0.76	0.76
11	1.16	0.87	0.83	0.85	0.86	0.80	0.79	0.83	0.85	0.84	0.81	0.80	0.80	0.81	0.81	0.81
12	1.26	0.92	0.88	0.90	0.91	0.84	0.84	0.89	0.90	0.89	0.86	0.85	0.85	0.86	0.86	0.86
13	1.30	0.92	0.89	0.91	0.92	0.87	0.85	0.92	0.91	0.89	0.87	0.86	0.86	0.88	0.88	0.88
14	1.34	0.93	0.90	0.93	0.93	0.88	0.87	0.94	0.92	0.90	0.88	0.88	0.87	0.90	0.89	0.89
15	1.36	0.93	0.91	0.94	0.93	0.89	0.88	0.95	0.92	0.90	0.89	0.88	0.88	0.91	0.90	0.90
16	1.38	0.93	0.91	0.95	0.93	0.90	0.88	0.97	0.93	0.91	0.90	0.89	0.89	0.91	0.91	0.91
17	1.39	0.94	0.92	0.95	0.94	0.90	0.89	0.98	0.93	0.91	0.90	0.90	0.89	0.92	0.91	0.91
18	1.40	0.94	0.92	0.95	0.94	0.91	0.90	0.98	0.93	0.91	0.91	0.90	0.89	0.92	0.92	0.92

RMSE - HICP - pre-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.14	0.13	0.14	0.14	0.13	0.14	0.14	0.14	0.14	0.16	0.14	0.14	0.14	0.13	0.13	0.14
2	0.21	0.20	0.21	0.22	0.20	0.21	0.21	0.20	0.21	0.21	0.20	0.20	0.20	0.21	0.21	0.21
3	0.27	0.25	0.26	0.26	0.25	0.26	0.26	0.26	0.26	0.25	0.25	0.25	0.25	0.25	0.25	0.25
4	0.30	0.27	0.29	0.29	0.28	0.28	0.29	0.29	0.29	0.27	0.28	0.28	0.28	0.28	0.28	0.28
5	0.33	0.29	0.31	0.31	0.29	0.30	0.31	0.31	0.32	0.29	0.30	0.30	0.30	0.30	0.30	0.30
6	0.34	0.30	0.32	0.32	0.30	0.31	0.31	0.32	0.33	0.29	0.30	0.30	0.30	0.31	0.31	0.31
7	0.36	0.30	0.32	0.32	0.31	0.33	0.32	0.33	0.34	0.29	0.30	0.30	0.30	0.31	0.31	0.31
8	0.38	0.31	0.33	0.33	0.32	0.34	0.33	0.33	0.35	0.29	0.31	0.31	0.30	0.32	0.32	0.32
9	0.38	0.31	0.33	0.34	0.31	0.34	0.33	0.33	0.33	0.28	0.30	0.30	0.29	0.32	0.32	0.32
10	0.37	0.29	0.34	0.34	0.30	0.34	0.35	0.35	0.34	0.28	0.30	0.30	0.29	0.33	0.33	0.33
11	0.35	0.29	0.34	0.34	0.29	0.34	0.35	0.37	0.34	0.28	0.30	0.30	0.29	0.33	0.33	0.33
12	0.37	0.30	0.36	0.36	0.30	0.36	0.37	0.39	0.34	0.28	0.31	0.30	0.30	0.35	0.35	0.35
13	0.36	0.30	0.35	0.34	0.30	0.40	0.35	0.38	0.30	0.27	0.29	0.28	0.27	0.34	0.34	0.34
14	0.34	0.30	0.34	0.33	0.30	0.40	0.34	0.37	0.29	0.27	0.29	0.27	0.26	0.34	0.34	0.33
15	0.34	0.30	0.34	0.33	0.30	0.40	0.33	0.36	0.29	0.27	0.28	0.27	0.26	0.33	0.33	0.33
16	0.33	0.30	0.34	0.32	0.30	0.39	0.31	0.37	0.29	0.27	0.28	0.26	0.25	0.32	0.32	0.32
17	0.32	0.30	0.31	0.32	0.31	0.39	0.29	0.38	0.29	0.26	0.27	0.25	0.24	0.31	0.31	0.31
18	0.33	0.30	0.29	0.31	0.30	0.39	0.29	0.39	0.28	0.26	0.26	0.25	0.24	0.31	0.30	0.30

RMSE - HICP - post-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.13	0.15	0.13	0.13	0.14	0.14	0.14	0.14	0.15	0.16	0.14	0.14	0.14	0.13	0.13	0.13
2	0.22	0.26	0.23	0.24	0.25	0.24	0.23	0.24	0.26	0.24	0.24	0.24	0.24	0.23	0.23	0.23
3	0.29	0.35	0.31	0.32	0.34	0.32	0.30	0.32	0.34	0.33	0.31	0.31	0.31	0.31	0.31	0.31
4	0.35	0.41	0.37	0.37	0.41	0.37	0.35	0.36	0.41	0.38	0.37	0.36	0.36	0.36	0.36	0.36
5	0.40	0.47	0.41	0.42	0.46	0.42	0.39	0.40	0.45	0.42	0.41	0.40	0.40	0.40	0.40	0.40
6	0.47	0.53	0.46	0.47	0.52	0.47	0.44	0.45	0.51	0.47	0.45	0.45	0.45	0.45	0.45	0.45
7	0.57	0.59	0.53	0.54	0.58	0.53	0.49	0.51	0.57	0.52	0.51	0.51	0.51	0.52	0.52	0.51
8	0.69	0.65	0.60	0.61	0.65	0.59	0.55	0.58	0.64	0.58	0.57	0.57	0.57	0.58	0.58	0.58
9	0.79	0.70	0.65	0.66	0.70	0.64	0.59	0.63	0.69	0.63	0.62	0.62	0.61	0.63	0.62	0.62
10	0.90	0.74	0.69	0.70	0.74	0.68	0.62	0.66	0.72	0.66	0.65	0.64	0.64	0.66	0.66	0.66
11	1.01	0.78	0.73	0.73	0.78	0.72	0.65	0.70	0.76	0.69	0.68	0.67	0.67	0.70	0.69	0.69
12	1.13	0.83	0.78	0.78	0.83	0.76	0.70	0.76	0.81	0.73	0.73	0.72	0.71	0.75	0.74	0.74
13	1.19	0.83	0.80	0.80	0.83	0.78	0.70	0.79	0.82	0.74	0.74	0.74	0.73	0.76	0.76	0.76
14	1.26	0.84	0.81	0.81	0.84	0.79	0.72	0.82	0.83	0.75	0.76	0.75	0.74	0.78	0.78	0.77
15	1.32	0.84	0.83	0.82	0.84	0.79	0.73	0.86	0.83	0.76	0.77	0.77	0.75	0.80	0.79	0.79
16	1.38	0.84	0.83	0.82	0.84	0.79	0.73	0.87	0.83	0.76	0.78	0.77	0.76	0.80	0.79	0.79
17	1.43	0.84	0.84	0.83	0.84	0.79	0.74	0.89	0.83	0.77	0.78	0.78	0.77	0.81	0.80	0.80
18	1.48	0.84	0.85	0.84	0.84	0.79	0.75	0.91	0.83	0.77	0.79	0.78	0.77	0.82	0.81	0.80

Note: Average 1: AR, VAR_C, VAR_U_sub, LSTAR, BVAR, individual equations; Average 2: AR, RIMA, LSTAR, BVAR, individual equations

RMSE - HICP excluding food and energy - full period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_ sub	VAR_C	Simple avg 1	MSE weight- ed avg1	Geom decl avg 1	Simple avg 2	MSE weight- ed avg2	Geom decl avg 2
1	0.09	0.10	0.08	0.08	0.10	0.10	0.08	0.09	0.11	0.10	0.08	0.08	0.08	0.08	0.08	0.08
2	0.10	0.14	0.09	0.09	0.14	0.11	0.09	0.10	0.15	0.13	0.10	0.10	0.10	0.09	0.09	0.09
3	0.12	0.20	0.11	0.11	0.19	0.13	0.11	0.12	0.20	0.19	0.13	0.12	0.12	0.11	0.11	0.11
4	0.14	0.25	0.13	0.13	0.25	0.15	0.13	0.14	0.24	0.24	0.15	0.13	0.13	0.12	0.12	0.12
5	0.17	0.31	0.15	0.15	0.30	0.16	0.15	0.15	0.29	0.29	0.17	0.15	0.15	0.14	0.14	0.13
6	0.20	0.37	0.18	0.17	0.36	0.18	0.17	0.18	0.34	0.35	0.20	0.18	0.17	0.16	0.16	0.15
7	0.23	0.43	0.21	0.20	0.42	0.21	0.21	0.20	0.39	0.41	0.24	0.20	0.20	0.18	0.18	0.18
8	0.27	0.48	0.24	0.23	0.48	0.23	0.24	0.22	0.45	0.46	0.27	0.23	0.22	0.21	0.21	0.20
9	0.31	0.54	0.27	0.26	0.54	0.25	0.27	0.25	0.50	0.52	0.31	0.26	0.25	0.24	0.23	0.23
10	0.34	0.60	0.30	0.28	0.60	0.27	0.30	0.27	0.55	0.58	0.34	0.28	0.27	0.26	0.25	0.25
11	0.38	0.66	0.33	0.31	0.65	0.30	0.33	0.29	0.61	0.63	0.37	0.31	0.29	0.28	0.28	0.27
12	0.42	0.72	0.37	0.34	0.71	0.32	0.36	0.32	0.66	0.69	0.40	0.34	0.32	0.31	0.30	0.29
13	0.43	0.72	0.38	0.34	0.72	0.33	0.37	0.32	0.67	0.70	0.41	0.34	0.32	0.31	0.30	0.29
14	0.45	0.73	0.40	0.36	0.73	0.34	0.40	0.33	0.68	0.70	0.42	0.36	0.34	0.33	0.32	0.31
15	0.47	0.73	0.41	0.37	0.73	0.35	0.42	0.34	0.69	0.71	0.43	0.37	0.35	0.34	0.34	0.32
16	0.48	0.74	0.43	0.38	0.74	0.36	0.43	0.35	0.70	0.70	0.44	0.38	0.36	0.35	0.35	0.33
17	0.50	0.74	0.44	0.39	0.74	0.37	0.45	0.36	0.71	0.70	0.45	0.39	0.38	0.37	0.36	0.34
18	0.50	0.74	0.45	0.40	0.74	0.38	0.46	0.37	0.71	0.70	0.45	0.40	0.38	0.37	0.37	0.35

RMSE - HICP excluding food and energy - pre-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_ sub	VAR_C	Simple avg 1	MSE weight- ed avg1	Geom decl avg 1	Simple avg 2	MSE weight- ed avg2	Geom Decl avg 2
1	0.07	0.08	0.06	0.06	0.08	0.10	0.06	0.07	0.09	0.08	0.07	0.07	0.07	0.06	0.06	0.06
2	0.09	0.12	0.08	0.08	0.11	0.12	0.07	0.09	0.13	0.11	0.09	0.08	0.08	0.08	0.08	0.08
3	0.11	0.16	0.10	0.10	0.16	0.13	0.09	0.11	0.17	0.15	0.11	0.10	0.10	0.10	0.09	0.09
4	0.14	0.21	0.12	0.12	0.20	0.15	0.10	0.13	0.21	0.19	0.13	0.12	0.12	0.11	0.11	0.11
5	0.17	0.26	0.14	0.14	0.25	0.16	0.13	0.16	0.25	0.24	0.15	0.14	0.13	0.13	0.13	0.12
6	0.20	0.31	0.17	0.17	0.31	0.17	0.15	0.18	0.30	0.28	0.18	0.16	0.16	0.15	0.15	0.14
7	0.24	0.36	0.20	0.20	0.36	0.19	0.18	0.21	0.34	0.33	0.20	0.19	0.19	0.17	0.17	0.17
8	0.27	0.42	0.23	0.23	0.41	0.21	0.21	0.23	0.39	0.38	0.24	0.22	0.21	0.20	0.20	0.19
9	0.30	0.47	0.26	0.25	0.47	0.21	0.23	0.24	0.44	0.43	0.26	0.24	0.23	0.21	0.21	0.20
10	0.34	0.53	0.29	0.27	0.52	0.23	0.26	0.25	0.49	0.48	0.28	0.26	0.25	0.23	0.23	0.22
11	0.38	0.58	0.32	0.29	0.58	0.24	0.29	0.26	0.54	0.52	0.31	0.28	0.27	0.25	0.25	0.24
12	0.41	0.64	0.35	0.32	0.63	0.26	0.33	0.28	0.59	0.57	0.34	0.32	0.31	0.28	0.27	0.27
13	0.43	0.65	0.36	0.32	0.65	0.27	0.34	0.27	0.60	0.59	0.34	0.32	0.31	0.28	0.27	0.27
14	0.44	0.66	0.37	0.33	0.66	0.29	0.35	0.26	0.61	0.59	0.35	0.33	0.32	0.29	0.28	0.28
15	0.45	0.66	0.38	0.32	0.67	0.31	0.36	0.25	0.62	0.59	0.36	0.33	0.32	0.29	0.28	0.28
16	0.46	0.67	0.38	0.33	0.67	0.33	0.37	0.24	0.63	0.59	0.36	0.33	0.33	0.30	0.29	0.28
17	0.47	0.68	0.39	0.33	0.68	0.36	0.39	0.25	0.64	0.59	0.37	0.34	0.34	0.31	0.30	0.29
18	0.47	0.69	0.40	0.33	0.69	0.38	0.40	0.25	0.64	0.59	0.38	0.34	0.34	0.31	0.30	0.30

RMSE - HICP excluding food and energy - post-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_ sub	VAR_C	Simple avg 1	MSE weight- ed avg1	Geom decl avg 1	Simple avg 2	MSE weight- ed avg2	Geom Decl avg 2
1	0.10	0.12	0.09	0.09	0.12	0.09	0.09	0.10	0.13	0.11	0.09	0.09	0.09	0.09	0.09	0.09
2	0.11	0.16	0.11	0.10	0.16	0.10	0.11	0.11	0.17	0.14	0.11	0.11	0.11	0.10	0.10	0.10
3	0.13	0.23	0.12	0.12	0.22	0.12	0.14	0.13	0.23	0.19	0.14	0.13	0.13	0.12	0.12	0.12
4	0.15	0.29	0.13	0.13	0.29	0.14	0.16	0.15	0.28	0.24	0.17	0.15	0.15	0.13	0.13	0.13
5	0.16	0.35	0.14	0.14	0.35	0.15	0.17	0.15	0.33	0.29	0.19	0.15	0.15	0.14	0.14	0.14
6	0.19	0.42	0.17	0.16	0.42	0.18	0.20	0.16	0.39	0.35	0.22	0.18	0.17	0.16	0.16	0.16
7	0.22	0.49	0.20	0.19	0.49	0.20	0.24	0.19	0.46	0.41	0.26	0.21	0.20	0.19	0.19	0.18
8	0.24	0.56	0.23	0.20	0.55	0.23	0.27	0.20	0.51	0.47	0.29	0.23	0.22	0.21	0.20	0.20
9	0.29	0.63	0.27	0.24	0.62	0.26	0.32	0.23	0.58	0.54	0.34	0.27	0.26	0.25	0.24	0.24
10	0.32	0.70	0.29	0.26	0.69	0.29	0.35	0.25	0.64	0.60	0.37	0.29	0.28	0.26	0.26	0.25
11	0.36	0.76	0.32	0.28	0.76	0.31	0.39	0.26	0.70	0.66	0.40	0.31	0.30	0.28	0.28	0.27
12	0.40	0.83	0.35	0.31	0.83	0.35	0.42	0.29	0.77	0.73	0.44	0.34	0.33	0.31	0.31	0.30
13	0.42	0.83	0.35	0.31	0.84	0.35	0.43	0.29	0.78	0.74	0.45	0.34	0.33	0.31	0.31	0.29
14	0.44	0.84	0.38	0.33	0.84	0.36	0.45	0.30	0.79	0.76	0.47	0.36	0.35	0.33	0.33	0.31
15	0.47	0.84	0.41	0.36	0.84	0.38	0.47	0.32	0.80	0.77	0.48	0.38	0.37	0.36	0.35	0.33
16	0.49	0.84	0.44	0.38	0.85	0.39	0.48	0.32	0.81	0.77	0.49	0.39	0.37	0.37	0.36	0.34
17	0.52	0.85	0.47	0.40	0.85	0.41	0.50	0.34	0.82	0.78	0.51	0.42	0.39	0.39	0.38	0.36
18	0.53	0.85	0.49	0.42	0.85	0.42	0.52	0.34	0.82	0.78	0.52	0.42	0.40	0.40	0.39	0.37

Note: Average 1: UM, AR, MA1, RIMA, RWAO, VAR_C, VAR_U_sub, LSTAR, BVAR, STIP; Average 2: AR, VAR_C

MAD - HICP - full period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom decl avg 2
1	0.13	0.13	0.12	0.12	0.12	0.13	0.12	0.13	0.13	0.14	0.12	0.12	0.12	0.12	0.12	0.12
2	0.20	0.21	0.20	0.20	0.20	0.19	0.20	0.20	0.20	0.21	0.20	0.20	0.19	0.20	0.20	0.20
3	0.28	0.29	0.27	0.27	0.28	0.27	0.27	0.27	0.28	0.27	0.26	0.26	0.26	0.26	0.26	0.26
4	0.34	0.34	0.33	0.32	0.34	0.32	0.32	0.32	0.34	0.33	0.32	0.32	0.32	0.32	0.32	0.32
5	0.40	0.39	0.37	0.37	0.39	0.37	0.36	0.36	0.38	0.37	0.36	0.36	0.36	0.36	0.36	0.36
6	0.47	0.43	0.41	0.41	0.43	0.41	0.39	0.40	0.42	0.41	0.40	0.40	0.40	0.40	0.40	0.40
7	0.54	0.47	0.46	0.46	0.47	0.45	0.43	0.45	0.46	0.45	0.44	0.44	0.43	0.44	0.44	0.44
8	0.62	0.51	0.51	0.51	0.51	0.50	0.48	0.50	0.50	0.48	0.48	0.48	0.48	0.49	0.49	0.49
9	0.70	0.55	0.55	0.55	0.55	0.54	0.52	0.54	0.53	0.52	0.52	0.52	0.52	0.54	0.53	0.53
10	0.76	0.58	0.59	0.60	0.58	0.58	0.56	0.59	0.57	0.55	0.56	0.55	0.55	0.57	0.57	0.57
11	0.83	0.62	0.62	0.63	0.62	0.61	0.59	0.63	0.61	0.59	0.59	0.59	0.58	0.61	0.60	0.60
12	0.91	0.66	0.66	0.67	0.66	0.65	0.63	0.69	0.66	0.63	0.63	0.63	0.62	0.65	0.65	0.65
13	0.95	0.66	0.67	0.67	0.66	0.67	0.64	0.71	0.66	0.63	0.64	0.64	0.63	0.66	0.65	0.65
14	0.97	0.67	0.68	0.68	0.67	0.68	0.64	0.72	0.66	0.64	0.65	0.64	0.63	0.67	0.66	0.66
15	0.99	0.67	0.68	0.69	0.67	0.69	0.64	0.72	0.67	0.64	0.66	0.65	0.64	0.67	0.67	0.67
16	0.99	0.68	0.69	0.70	0.68	0.69	0.64	0.74	0.67	0.64	0.66	0.65	0.64	0.68	0.67	0.67
17	1.01	0.68	0.69	0.70	0.68	0.70	0.65	0.74	0.68	0.65	0.66	0.66	0.65	0.69	0.68	0.68
18	1.02	0.68	0.69	0.71	0.68	0.70	0.65	0.75	0.68	0.65	0.66	0.66	0.65	0.69	0.68	0.68

MAD - HICP - pre-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.11	0.10	0.10	0.11	0.10	0.11	0.11	0.11	0.11	0.13	0.11	0.11	0.11	0.10	0.10	0.10
2	0.16	0.15	0.17	0.18	0.15	0.16	0.17	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16	0.16
3	0.21	0.19	0.21	0.22	0.19	0.21	0.21	0.20	0.21	0.20	0.20	0.20	0.20	0.20	0.20	0.20
4	0.25	0.22	0.24	0.25	0.22	0.23	0.24	0.24	0.25	0.22	0.23	0.23	0.23	0.24	0.24	0.24
5	0.27	0.24	0.26	0.26	0.24	0.25	0.25	0.26	0.26	0.23	0.24	0.24	0.24	0.25	0.25	0.25
6	0.28	0.24	0.27	0.26	0.24	0.26	0.26	0.26	0.27	0.23	0.25	0.25	0.25	0.26	0.26	0.26
7	0.30	0.24	0.27	0.27	0.25	0.27	0.26	0.27	0.26	0.22	0.24	0.24	0.24	0.25	0.25	0.25
8	0.31	0.24	0.27	0.26	0.25	0.28	0.26	0.27	0.26	0.22	0.24	0.24	0.24	0.26	0.26	0.26
9	0.32	0.24	0.27	0.27	0.24	0.29	0.27	0.27	0.25	0.22	0.25	0.24	0.24	0.26	0.26	0.26
10	0.31	0.23	0.27	0.27	0.24	0.28	0.28	0.29	0.25	0.22	0.24	0.24	0.23	0.26	0.26	0.26
11	0.29	0.23	0.27	0.27	0.24	0.28	0.28	0.32	0.27	0.23	0.24	0.25	0.24	0.26	0.26	0.26
12	0.31	0.24	0.28	0.28	0.25	0.30	0.29	0.34	0.28	0.25	0.25	0.25	0.25	0.27	0.27	0.27
13	0.30	0.25	0.27	0.26	0.25	0.31	0.27	0.33	0.25	0.23	0.24	0.24	0.23	0.27	0.27	0.27
14	0.30	0.25	0.27	0.27	0.25	0.32	0.26	0.31	0.24	0.22	0.24	0.24	0.23	0.26	0.26	0.25
15	0.28	0.25	0.27	0.27	0.25	0.33	0.25	0.30	0.25	0.23	0.24	0.24	0.23	0.26	0.26	0.26
16	0.26	0.25	0.27	0.26	0.25	0.34	0.24	0.31	0.24	0.23	0.23	0.23	0.22	0.25	0.25	0.25
17	0.27	0.25	0.25	0.26	0.25	0.34	0.23	0.32	0.24	0.23	0.22	0.22	0.21	0.26	0.26	0.25
18	0.26	0.25	0.24	0.25	0.25	0.33	0.23	0.32	0.23	0.22	0.21	0.21	0.20	0.25	0.24	0.24

MAD - HICP - post-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.10	0.12	0.10	0.10	0.11	0.11	0.10	0.11	0.11	0.12	0.10	0.10	0.10	0.10	0.10	0.10
2	0.17	0.22	0.19	0.19	0.21	0.19	0.19	0.20	0.21	0.21	0.19	0.19	0.19	0.19	0.19	0.19
3	0.24	0.30	0.26	0.27	0.30	0.27	0.26	0.27	0.30	0.28	0.27	0.27	0.27	0.26	0.26	0.26
4	0.28	0.35	0.31	0.31	0.35	0.31	0.31	0.31	0.35	0.33	0.32	0.31	0.31	0.31	0.31	0.31
5	0.32	0.39	0.34	0.35	0.39	0.35	0.33	0.34	0.38	0.35	0.34	0.34	0.34	0.34	0.34	0.33
6	0.39	0.42	0.37	0.38	0.42	0.38	0.35	0.36	0.41	0.38	0.37	0.37	0.36	0.36	0.36	0.36
7	0.48	0.46	0.44	0.44	0.46	0.43	0.40	0.43	0.45	0.42	0.41	0.41	0.41	0.42	0.42	0.42
8	0.58	0.51	0.51	0.51	0.51	0.48	0.46	0.50	0.50	0.46	0.47	0.46	0.46	0.49	0.49	0.49
9	0.69	0.55	0.56	0.56	0.55	0.53	0.50	0.55	0.55	0.50	0.51	0.51	0.51	0.54	0.53	0.53
10	0.78	0.59	0.60	0.60	0.58	0.58	0.54	0.59	0.57	0.52	0.54	0.54	0.54	0.58	0.57	0.57
11	0.88	0.63	0.65	0.64	0.62	0.62	0.58	0.63	0.60	0.55	0.59	0.58	0.58	0.62	0.62	0.62
12	1.00	0.67	0.69	0.68	0.66	0.66	0.61	0.68	0.64	0.58	0.63	0.62	0.61	0.66	0.66	0.66
13	1.06	0.67	0.70	0.69	0.67	0.68	0.62	0.71	0.66	0.59	0.64	0.63	0.62	0.67	0.67	0.66
14	1.12	0.67	0.71	0.70	0.67	0.68	0.62	0.73	0.67	0.60	0.65	0.64	0.63	0.68	0.68	0.68
15	1.17	0.67	0.72	0.70	0.67	0.68	0.63	0.76	0.67	0.61	0.66	0.65	0.65	0.69	0.69	0.69
16	1.21	0.67	0.73	0.71	0.67	0.68	0.63	0.78	0.67	0.61	0.67	0.66	0.65	0.70	0.69	0.69
17	1.24	0.67	0.73	0.72	0.67	0.68	0.63	0.78	0.67	0.61	0.67	0.66	0.65	0.70	0.70	0.69
18	1.27	0.68	0.74	0.73	0.68	0.68	0.64	0.80	0.67	0.61	0.67	0.66	0.65	0.71	0.70	0.70

Note: Average 1: UM, AR, MA1, RIMA, RWAO, VAR_C, VAR_U_sub, LSTAR, BVAR, STIP; Average 2: AR, VAR_C

MAD - HICP excluding food and energy - full period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom decl avg 2
1	0.06	0.08	0.06	0.05	0.07	0.07	0.06	0.06	0.08	0.07	0.06	0.06	0.06	0.06	0.06	0.06
2	0.08	0.12	0.07	0.07	0.12	0.08	0.07	0.08	0.12	0.11	0.08	0.07	0.07	0.07	0.07	0.07
3	0.10	0.17	0.09	0.09	0.16	0.10	0.09	0.09	0.16	0.15	0.10	0.09	0.09	0.08	0.08	0.08
4	0.11	0.22	0.10	0.10	0.21	0.12	0.10	0.11	0.20	0.20	0.12	0.11	0.10	0.10	0.10	0.09
5	0.14	0.27	0.12	0.12	0.27	0.13	0.12	0.12	0.24	0.25	0.14	0.12	0.12	0.11	0.11	0.11
6	0.16	0.32	0.15	0.14	0.32	0.15	0.14	0.14	0.29	0.30	0.17	0.15	0.14	0.13	0.13	0.13
7	0.19	0.38	0.17	0.16	0.37	0.17	0.17	0.15	0.34	0.36	0.20	0.17	0.16	0.15	0.15	0.15
8	0.22	0.43	0.20	0.19	0.43	0.19	0.20	0.17	0.39	0.41	0.23	0.19	0.18	0.17	0.17	0.17
9	0.25	0.49	0.23	0.21	0.48	0.21	0.22	0.19	0.44	0.46	0.25	0.22	0.21	0.20	0.19	0.19
10	0.28	0.54	0.25	0.23	0.54	0.22	0.24	0.22	0.49	0.51	0.28	0.23	0.22	0.21	0.21	0.20
11	0.31	0.60	0.27	0.25	0.59	0.24	0.27	0.23	0.54	0.56	0.30	0.25	0.24	0.23	0.23	0.22
12	0.35	0.65	0.30	0.27	0.65	0.26	0.30	0.26	0.59	0.61	0.33	0.28	0.26	0.25	0.25	0.24
13	0.36	0.66	0.31	0.28	0.66	0.27	0.30	0.26	0.60	0.62	0.33	0.28	0.27	0.25	0.25	0.24
14	0.38	0.66	0.32	0.29	0.66	0.28	0.32	0.27	0.62	0.63	0.34	0.29	0.28	0.27	0.27	0.25
15	0.40	0.67	0.33	0.30	0.67	0.29	0.34	0.28	0.63	0.63	0.35	0.30	0.29	0.28	0.28	0.26
16	0.41	0.67	0.34	0.31	0.67	0.29	0.34	0.30	0.63	0.63	0.35	0.31	0.29	0.28	0.29	0.27
17	0.42	0.68	0.35	0.32	0.68	0.30	0.36	0.31	0.64	0.63	0.36	0.32	0.31	0.30	0.30	0.28
18	0.43	0.68	0.36	0.32	0.68	0.31	0.37	0.31	0.65	0.63	0.37	0.33	0.31	0.30	0.30	0.28

MAD - HICP excluding food and energy - pre-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.05	0.06	0.05	0.05	0.06	0.07	0.05	0.06	0.07	0.06	0.05	0.05	0.05	0.05	0.05	0.05
2	0.07	0.10	0.06	0.06	0.10	0.09	0.06	0.08	0.10	0.09	0.07	0.06	0.06	0.06	0.06	0.06
3	0.09	0.14	0.08	0.08	0.13	0.10	0.07	0.09	0.13	0.13	0.08	0.08	0.08	0.08	0.08	0.08
4	0.11	0.18	0.10	0.10	0.17	0.11	0.08	0.11	0.16	0.16	0.10	0.09	0.09	0.09	0.09	0.09
5	0.14	0.23	0.12	0.12	0.22	0.13	0.10	0.12	0.21	0.21	0.12	0.12	0.11	0.11	0.11	0.11
6	0.17	0.28	0.14	0.14	0.27	0.14	0.12	0.14	0.25	0.26	0.15	0.14	0.14	0.13	0.13	0.12
7	0.19	0.33	0.17	0.16	0.32	0.16	0.15	0.15	0.29	0.30	0.17	0.17	0.16	0.15	0.15	0.14
8	0.23	0.38	0.20	0.19	0.37	0.18	0.17	0.18	0.34	0.35	0.20	0.19	0.19	0.17	0.17	0.17
9	0.26	0.44	0.22	0.21	0.43	0.19	0.19	0.18	0.39	0.40	0.22	0.20	0.20	0.19	0.18	0.18
10	0.28	0.49	0.24	0.23	0.48	0.20	0.21	0.19	0.44	0.45	0.24	0.22	0.21	0.20	0.20	0.19
11	0.32	0.55	0.27	0.24	0.54	0.21	0.24	0.20	0.48	0.50	0.27	0.24	0.23	0.21	0.21	0.20
12	0.35	0.60	0.29	0.26	0.60	0.23	0.27	0.22	0.54	0.54	0.30	0.26	0.25	0.23	0.23	0.22
13	0.36	0.61	0.30	0.26	0.61	0.25	0.28	0.20	0.56	0.56	0.30	0.27	0.26	0.23	0.23	0.23
14	0.37	0.62	0.31	0.27	0.62	0.26	0.28	0.20	0.57	0.56	0.31	0.27	0.27	0.24	0.24	0.24
15	0.38	0.63	0.31	0.27	0.63	0.26	0.30	0.20	0.58	0.57	0.31	0.28	0.28	0.25	0.25	0.24
16	0.38	0.64	0.32	0.27	0.64	0.27	0.31	0.20	0.59	0.57	0.32	0.29	0.28	0.25	0.25	0.25
17	0.40	0.65	0.32	0.28	0.65	0.28	0.32	0.20	0.60	0.57	0.33	0.30	0.29	0.26	0.26	0.25
18	0.40	0.65	0.33	0.28	0.65	0.28	0.34	0.21	0.61	0.57	0.34	0.30	0.30	0.26	0.26	0.25

MAD - HICP excluding food and energy - post-crisis period

	RWAO	UM	AR	RIMA	MA1	LSTAR	Ind. equ.	BVAR	VAR_U_sub	VAR_C	Simple avg 1	MSE weight-ed avg1	Geom decl avg 1	Simple avg 2	MSE weight-ed avg2	Geom Decl avg 2
1	0.07	0.09	0.06	0.06	0.09	0.07	0.07	0.07	0.09	0.08	0.07	0.07	0.07	0.06	0.06	0.06
2	0.08	0.14	0.08	0.07	0.14	0.08	0.09	0.08	0.14	0.11	0.09	0.08	0.08	0.08	0.08	0.08
3	0.10	0.20	0.09	0.09	0.20	0.10	0.10	0.09	0.19	0.15	0.11	0.10	0.10	0.09	0.09	0.09
4	0.11	0.26	0.11	0.10	0.26	0.12	0.12	0.11	0.24	0.20	0.14	0.11	0.11	0.10	0.10	0.10
5	0.13	0.32	0.11	0.11	0.32	0.13	0.14	0.11	0.29	0.25	0.16	0.13	0.12	0.11	0.11	0.11
6	0.15	0.39	0.14	0.13	0.38	0.14	0.16	0.12	0.35	0.31	0.18	0.14	0.14	0.12	0.12	0.12
7	0.18	0.45	0.16	0.15	0.45	0.16	0.20	0.14	0.42	0.37	0.22	0.17	0.16	0.15	0.15	0.14
8	0.20	0.52	0.19	0.17	0.51	0.18	0.23	0.15	0.47	0.42	0.24	0.18	0.18	0.17	0.16	0.16
9	0.24	0.58	0.22	0.19	0.58	0.22	0.27	0.18	0.53	0.49	0.28	0.22	0.21	0.20	0.20	0.19
10	0.27	0.65	0.24	0.21	0.64	0.24	0.29	0.20	0.59	0.55	0.31	0.23	0.22	0.21	0.21	0.20
11	0.31	0.71	0.27	0.23	0.71	0.26	0.32	0.21	0.65	0.61	0.34	0.26	0.25	0.23	0.23	0.22
12	0.35	0.78	0.29	0.25	0.78	0.29	0.35	0.24	0.72	0.67	0.38	0.28	0.27	0.25	0.25	0.24
13	0.37	0.78	0.29	0.26	0.78	0.29	0.36	0.25	0.73	0.68	0.37	0.28	0.27	0.26	0.25	0.24
14	0.39	0.79	0.31	0.28	0.79	0.30	0.37	0.26	0.74	0.70	0.38	0.29	0.28	0.27	0.26	0.25
15	0.42	0.79	0.34	0.30	0.79	0.31	0.39	0.28	0.75	0.71	0.40	0.30	0.29	0.28	0.28	0.27
16	0.44	0.79	0.35	0.31	0.79	0.32	0.40	0.29	0.76	0.71	0.41	0.32	0.30	0.29	0.30	0.28
17	0.46	0.80	0.38	0.33	0.80	0.34	0.42	0.30	0.77	0.71	0.42	0.34	0.32	0.32	0.31	0.30
18	0.48	0.80	0.39	0.35	0.80	0.35	0.43	0.31	0.77	0.72	0.43	0.35	0.33	0.33	0.33	0.31

Note: Average 1: UM, AR, MA1, RIMA, RWAO, VAR_C, VAR_U_sub, LSTAR, BVAR, STIP; Average 2: AR, VAR_C

Appendix II Details on forecasting methods for inflation

1. Random Walk Models

The random walk models include the random walk with drift and the random walk model according to Atkenson and Ohanian (2001).

2. AR Models

The lag selection is done recursively based on the Bayesian Information Criterion. The number of lags from which these criteria select for the inflation models ranges from 0 and 6 months.

3. STAR Models

The Smooth Transition Autoregressive (STAR) Model is included in the analysis. This model is a generalisation of the 2-regime threshold autoregressive model, that allows the estimation of the – potentially smooth - transition from one regime to the other and can help forecasting variables with asymmetric behaviour (see e.g. Teraesvirta, 2006).

4. VAR Models

Two VAR model versions are being employed to generate forecasts for all variables contained in the models. The first model comprises HICP components but not the aggregate itself (i.e. 5 variables: processed food, unprocessed food, industrial goods, services, energy) and the aggregate, but does not include any exogenous variables on which the forecasts could be conditioned. The second model does not include any of the other HICP components, but uses the producer prices of consumer goods, a moving average of interpolated monthly unit labour costs, oil prices, non-oil commodity prices and the nominal effective exchange rate as explanatory variables. These variables are endogenous in the estimation of the VAR, but the forecasts are conditioned on the assumptions as they are used in the Eurosystem Macroeconomic Projection Exercises.

All HICP index variables are modelled in month-on-month log-differences, i.e. approximate percentage changes. Similar to the AR models, the lag selection is done recursively based on the Bayesian Information Criterion. The number of lags from which these criteria select for the inflation models ranges from 0 and 6 months.

5. MA(1)

This model is employed to approximate the unobserved component model by Stock and Watson (2007) as these authors suggest.

6. Individual equation framework for inflation:

The individual equations for inflation (individual equation framework) are an updated version of Benalal et al. (2004) and are also described in ECB (2010). According to these equations, unprocessed food prices are modelled as an AR model with 12 lags with a seasonal dummy for January; the model for processed food prices includes food commodity prices in euro (lag 9), tobacco taxes, a 3 month moving average of unit labour costs (lag 5) and 2 lags of the dependent variable; non-energy industrial goods prices are explained by the effective exchange rate (lag 1), the VAT, GDP, consumer goods producer prices (lag 5) and lags 5 and 6 of the dependent variable; energy prices have as explanatory variables oil prices in USD and the USD exchange rate (both lags 0 and 1), energy taxes and seasonal dummies for January and November; services prices are modelled with a 3 month moving average of compensation per employee (lags 3 and 9), consumer goods producer prices (lags 5 and 7) and lags 1 to 5 of the dependent variable; the direct approach for headline HICP uses a model with a 3 month moving average of unit labour costs, a 3 month moving average of oil prices in euro (lags 0 and 3), the nominal effective exchange rate (lags 0 to 0), consumer goods producer prices (lags 0 and 1) and lags 1 to 12 of the dependent variable; the direct equation for the HICP excluding food and energy uses a moving average of unit labour costs, the nominal effective exchange rate (lag 7), consumer goods producer prices (lags 0 and 1) and lags 2, 5 and 6 of the dependent variable. Consumer goods producer prices, which are used as explanatory variable in some equations, are modelled with non-oil commodity prices (lag 6), the nominal effective exchange rate (lag 4), GDP and lags 1 to 4 of the dependent variable. All equations are estimated in dlogs on seasonal adjusted prices.

6. BVAR

The BVAR model is the one proposed by Giannone et al. (2014), with priors following Giannone, Lenza and Primiceri (2015). It allows for all possible interactions between main components of HICP and their determinants and captures indirect and second-round effects.

Their dataset is very close to the one for the individual equation approach, it is of monthly frequency, and includes 14 variables. Namely, the 5 components of HICP, PPI consumer goods, unit labour costs, GDP, compensation per employee, oil price in US dollars, food commodity prices, commodity prices excluding food, EUR/USD exchange rate and nominal effective exchange rate.

In order to mimic the information available to forecasters in real time, they gathered, to the extent possible, 39 vintages of the real-time data available for the corresponding quarterly Eurosystem/ECB staff projection exercises from March 2000 to September 2009.[17] They reconstruct the availability of official data, as well as the assumptions on future paths that were available in real time for almost all the variables which are used to condition the inflation forecast in the individual equation and the conditional BVAR. For the quarterly variables (unit labour costs, GDP and compensation per employee), they make use of the historical data, as well as the, publicly available, projected path of annual GDP growth rates. For what concerns unit labour costs and compensation per employee, no particular future projected path is assumed. Oil and non-oil commodity prices are based on Bloomberg data on spot and futures prices. As the back-data were not complete for all of the 39 vintages, they reconstructed the oil and non-oil commodities real-time database on the basis of spot and futures' prices available at earlier points

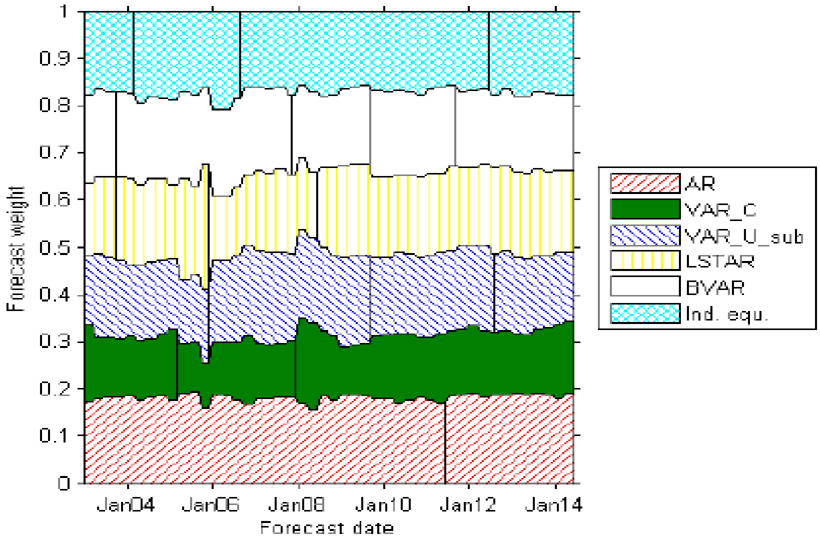
[17] Note that Eurosystem projections were at the time published in form of ranges. In order to obtain point estimates, they took the mid-point of the ranges.

in time. Finally, exchange rates are unrevised and the assumed future path in the forecasting exercises in their case is actually flat at the latest observed value at each different data cut-off point of the different quarterly projections (implied random walk assumption). Notice that taxes also enter the original individual equations. Accordingly, they make use of them in the study of the forecasting performance of the individual equations. However, in the BVARs, they have chosen to exclude them since they were not found to improve on forecast accuracy.

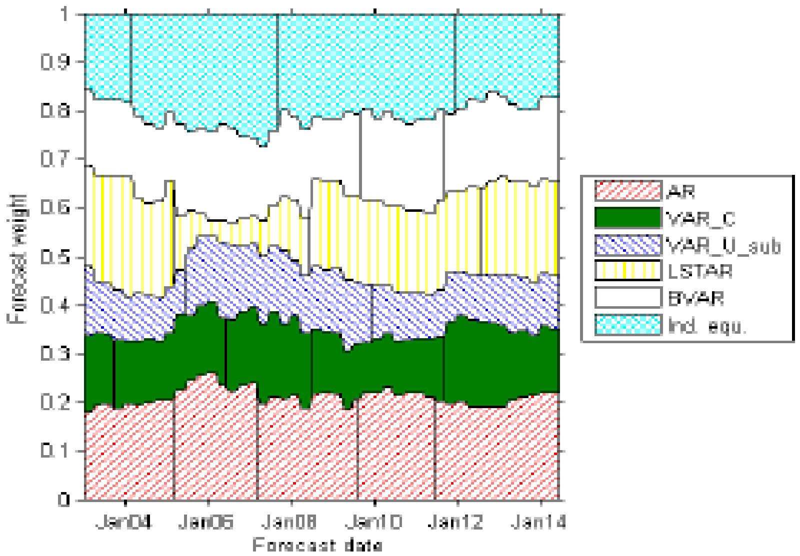
Appendix III Weight charts with Bai-Perron break dates

Weights with Bai-Perron Break dates (vertical lines), Combination 2, full forecast sample

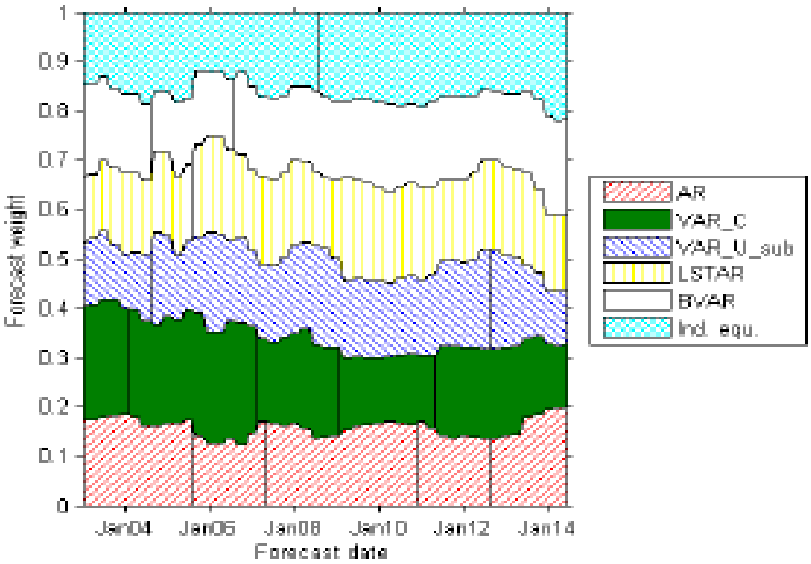
Horizon 1, Headline inflation (hlagg)



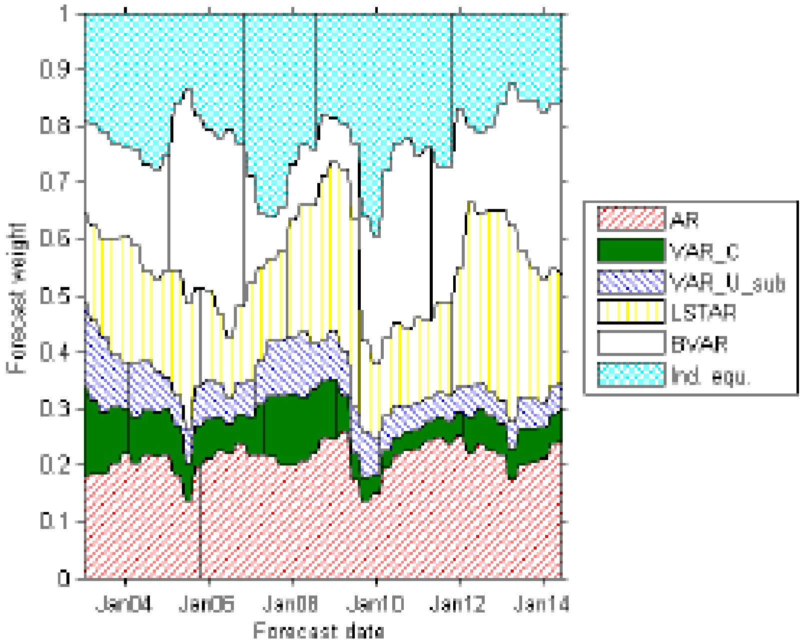
Horizon 1, Inflation excluding energy and food (exagg)



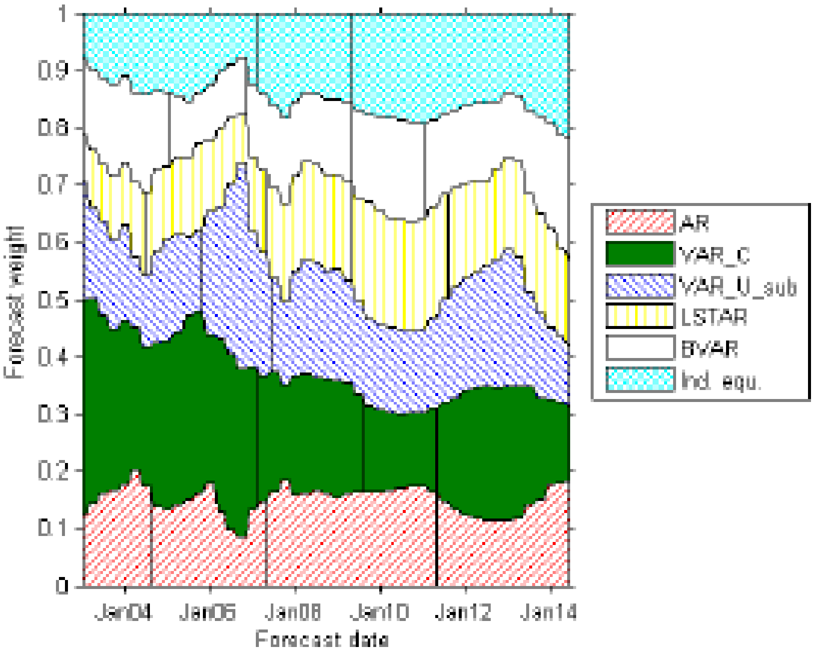
Horizon 6, Headline inflatioin (hlagg)



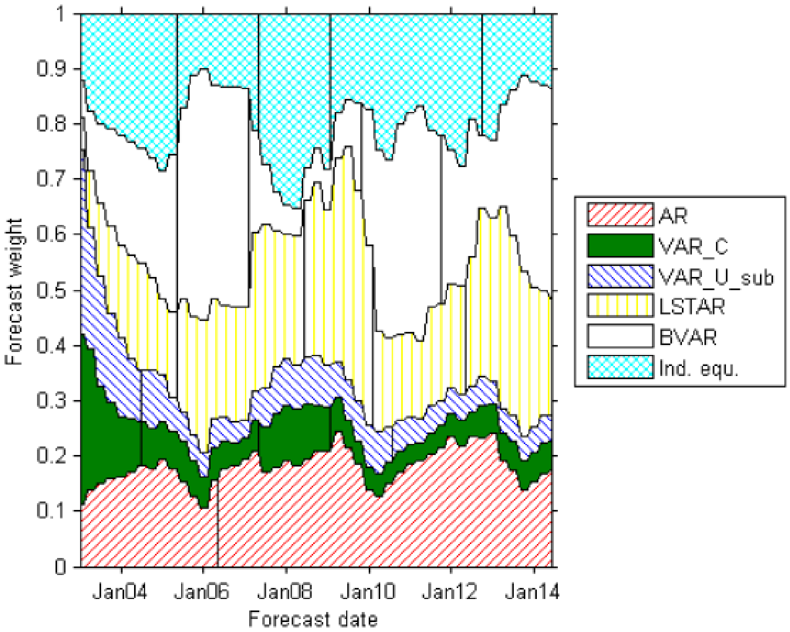
Horizon 6, Inflation excluding energy and food (exagg)



Horizon 12, Headline inflation (hlagg)



Horizon 12, Inflation excluding energy and food (exagg)



Appendix IV Forecasting the aggregate versus combining component forecasts

As mentioned in Section 2.1, a combination issue separate from the one just discussed that focuses on the combination of forecasts of the same variable of interest, is whether combining forecasts of different disaggregate variables to forecast the aggregate ("indirect forecast" of the aggregate) is preferable to combining disaggregate information in a model for the aggregate and use this for forecasting ("direct forecast" of the aggregate).

This has been investigated empirically for the euro area by Benalal, Diaz Del Hoyo, Landau, Roma and Skudelny (2004), and Hubrich (2005), and has been analysed theoretically, by simulation and empirically for the US in Hendry and Hubrich (2011) Benalal, Diaz Del Hoyo, Landau, Roma and Skudelny (2004) find that the direct forecast of the aggregate provides better results than the aggregation of component forecasts for 12 and 18 months ahead for overall HICP while for the HICP excluding unprocessed food and energy, the forecast performance was better for the aggregation of the component forecast. Hubrich (2005) also found for 12-months ahead forecasts that the direct forecast of the aggregate outperformed the aggregation of component forecasts for total HICP, but not for the HICP excluding unprocessed food and energy.

Hendry and Hubrich (2011) derive analytical results and conclude with two main findings: (1) Unknown breaks at the forecast origin (location shifts or slope parameter changes) do not affect the *relative forecast accuracy* of forecasting the aggregate directly or aggregating component forecasts. This is in contrast to the findings in the usual forecast combination literature focusing on combining forecasts of the same variable. Of course, the absolute forecast accuracy might be affected substantially. (2) Slope parameter misspecification and estimation uncertainty are the main sources of relative forecast error differences.]

Hendry and Hubrich (2011) recommend reducing estimation uncertainty by (a) variable selection procedures (or Bayesian shrinkage) and (b) methods to combine disaggregate information (e.g. factor models or Bayesian shrinkage). The Bayesian VAR model included in this study addresses the issue of estimation uncertainty / curse of dimensionality by imposing prior information on the VARs for forecasting the disaggregates.

We present some results on forecasting the aggregate versus aggregating component forecasts in the context of the models employed in this study below. In addition to comparing the indirect forecasts, i.e. the aggregation of the component forecasts, described in the main text, we also computed and compared the results with the direct forecasts of headline HICP inflation and of HICP excluding food and energy inflation for the different models.We describe the details of the forecast experiments and discuss the results.

For the individual equation framework the direct equation for headline inflation contains up to 12 lags of the dependent variable, a moving average of unit labour costs (lags 0 to 3), a moving average of oil prices (lags 0 to 3), the nominal effective exchange rate (lags 0 to 9) and the PPI consumer goods (lags 0 and 1). For the HICP excluding food and energy, the equation includes lags 2, 5 and 6 of the dependent variable, a moving average of unit labour costs, the nominal effective exchange rate (lag 7) and the PPI consumer goods (lags 0 and 1). For the other models, the same variables are used for the direct approach as for the components based forecasts.

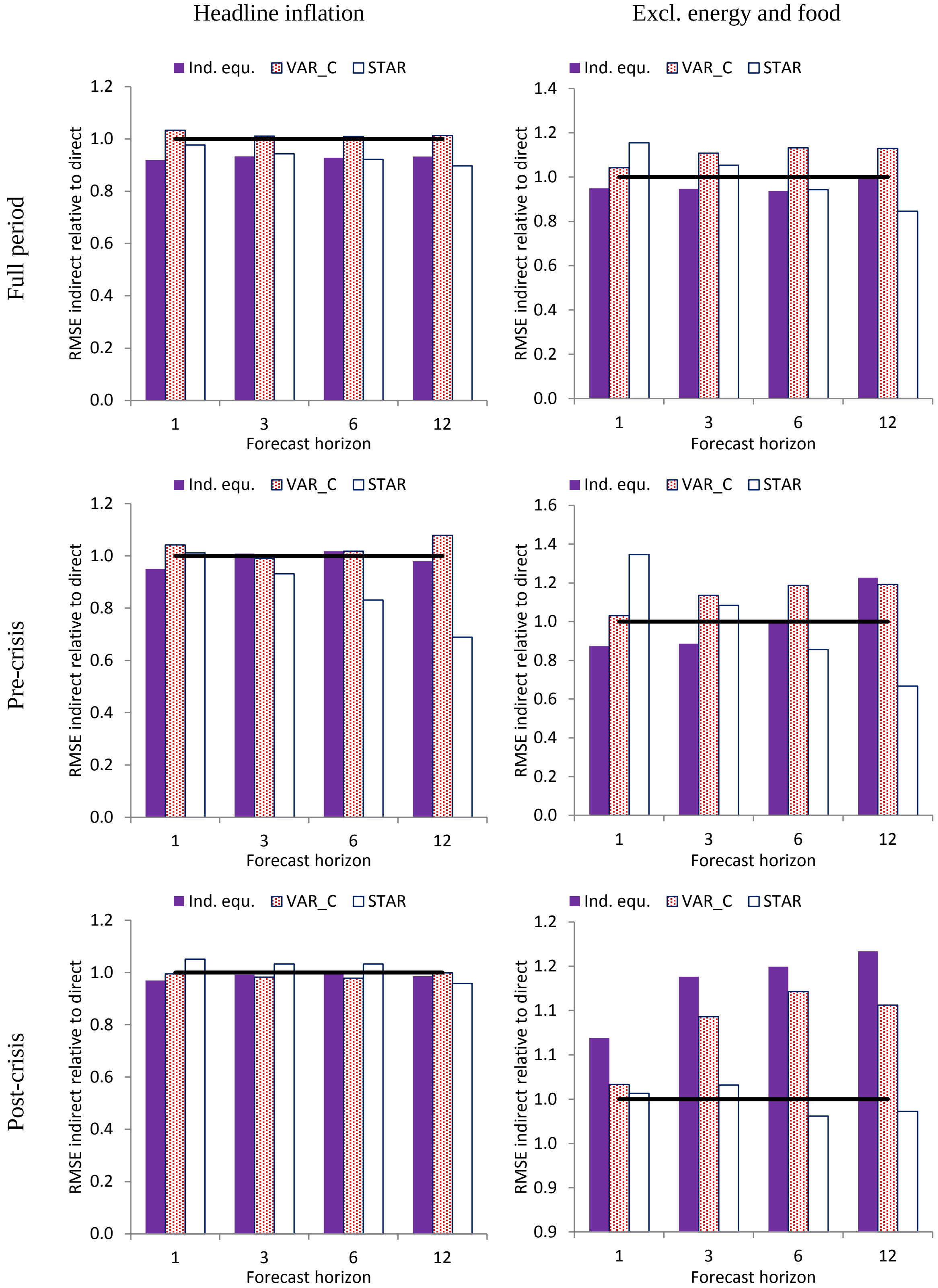
For the AR models, the direct AR forecast is always clearly better than the indirect forecast of aggregate inflation for the pre-crisis period, while for the period including the crisis period it clearly outperforms the indirect forecast only from a horizon of 7 months onwards and is very similar to it before. For the HICP excluding food and energy, the indirect forecast performs somewhat better. However, as we want to have a consistent benchmark across models and aggregate, our comparisons above were all based on the direct AR forecast.

Chart 9 shows the RMSE of the indirect (bottom-up) approach relative to the direct approach for headline inflation, both for the pre-crisis period and the period including the crisis. A number below one indicates that the indirect approach has a better forecast performance than the direct approach.

The results are quite different across models. For the individual equation framework, the indirect approach works somewhat better than the direct approach for the full sample for headline inflation and, though only up to 6 months ahead, for core inflation. In the two sub-periods, the results of the two approaches are very similar for headline inflation while for core inflation the indirect approach works better only for short horizons in the pre-crisis period and cannot outperform the AR for the post-crisis period. The VAR_C model provides similar or better results when using the direct rather than the indirect approach for all horizons, with larger differences for core inflation. The STAR model yields similar or better results for the indirect approach in particular over the longer run. This is an interesting result since Benalal et al. (2004) and Hubrich (2005) who find that indirect approaches tend to outperform direct approaches in particular over the short run using linear models. The results above show that different reference periods can lead to different results regarding whether it is better to forecast components or aggregates.

We also compared the direct and indirect forecasts of the two different inflation aggregates we consider in this paper for the different combination schemes (see Chart 10). When looking at the results for the combination models, we find that forecast accuracy is very similar between the direct and the indirect approach, both for headline and core inflation, although for core inflation there is some improvement of the indirect over the direct forecast. The difference between the periods is rather marginal.

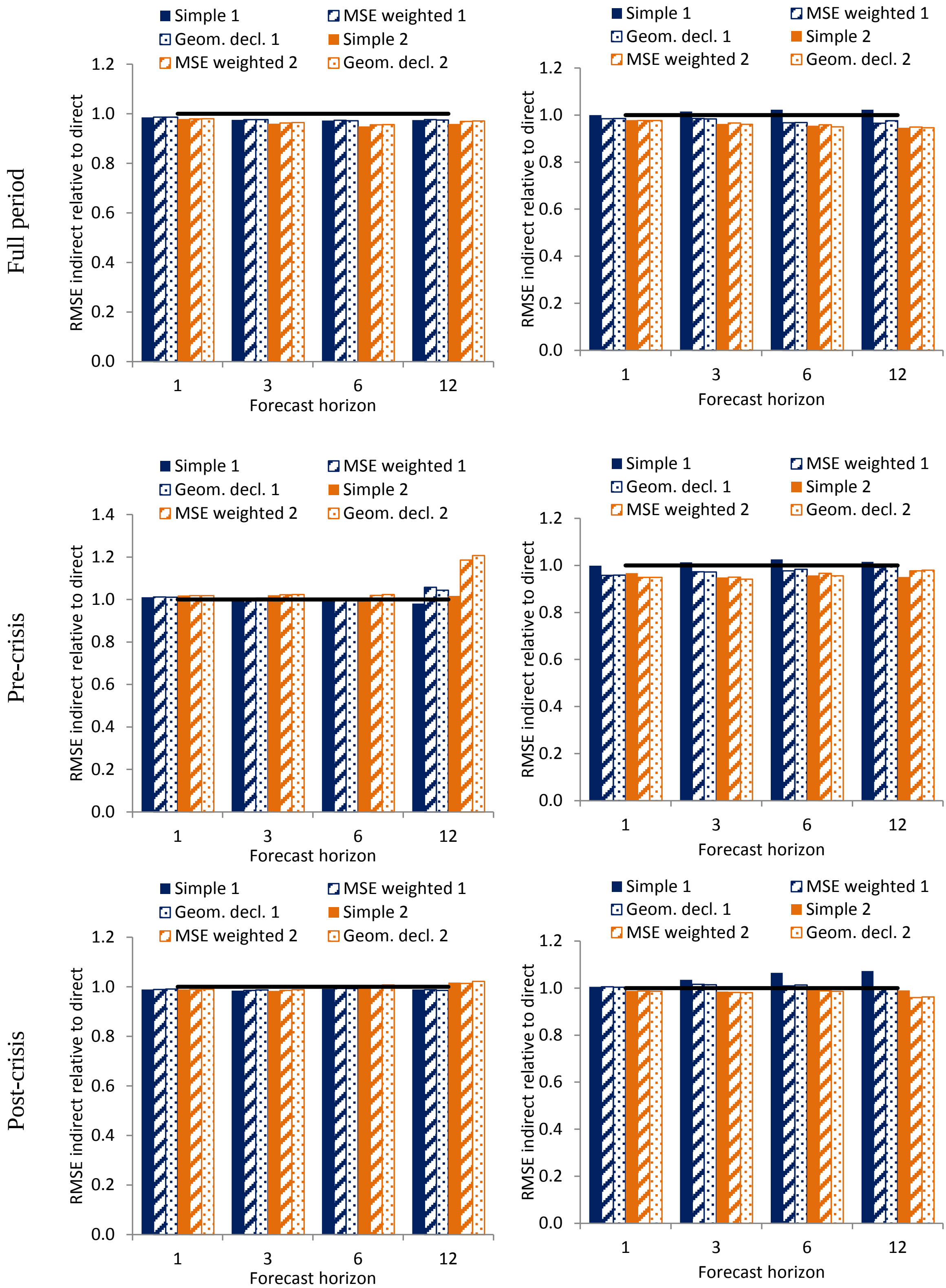
Chart 9 Indirect RMSE relative to direct RMSE – individual models



Note: "indirect forecast" of the aggregate (and the associated RMSFE) refers to combining forecasts of different disaggregate variables to forecast the aggregate; "direct forecast" of the aggregate (and its RMSFE) refers to the aggregate forecasting model.

Chart 10 Indirect RMSE relative to direct RMSE – combinations

Headline inflation | Excl. energy and food



Acknowledgements

We thank Tim van Ark, Marco Buchmann and Paul Körbitz for excellent research assistance and Domenico Giannone and Michele Lenza for sharing with us the codes for the Bayesian Vector Autoregressive Model. We thank Hans-Joachim Klöckers, Julian Morgan, John Roberts and participants of ESCB workshops for comments. The views expressed in this paper are those of the authors and do not necessarily reflect those of the European Central Bank or the Federal Reserve Board. The first draft of the paper was circulated in July 2010 when the first author was working at the European Central Bank.

Kirstin Hubrich
Federal Reserve Board; email: kirstin.hubrich@frb.gov

Frauke Skudelny
European Central Bank; email: frauke.skudelny@ecb.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

ISSN 1725-2806 (pdf) DOI 10.2866/052798 (pdf)
ISBN 978-92-899-2220-3 (pdf) EU catalogue No QB-AR-16-089-EN-N (pdf)