Weigh(t)ing the Basket: Aggregate and Component-Based Inflation Forecasts for the Euro Area

Jakub Chalmovianský[1] Mario Porqueddu[2,3] Andrej Sokol[4,5]

[1]Masaryk University
[2]European Central Bank
[3]Bank of Italy
[4]Bloomberg Economics
[5]Centre for Macroeconomics

EEA 2022, 25 August 2022

Disclaimer: the views expressed in this paper are those of the authors, and do not necessarily reflect those of the European Central Bank, the Bank of Italy or Bloomberg.

This paper is published as ECB WP No 2501 / December 2020

Introduction

- Headline HICP and HICP excl. energy and food (HICPX) can be represented as weighted sums of their components
- *Are top-down forecasts of HICP(X) better than aggregated (bottom-up) forecasts of its components?*
- **Theoretical results**: bottom-up approach can improve overall forecast performance compared to the top-down approach (under the assumption of known DGP, as in Theil (1954) and Lütkepohl (1984)).
- In practice DGP is unknown and estimated, forecast accuracy depends on the true DGP $\Rightarrow$ an **empirical issue** (see Lütkepohl (1987)).
- Empirical literature shows no clear-cut results
- Important for policy (e.g. how to design the forecasting process)

Empirical Literature

- Empirical studies:
  - aggregation of component forecasts: Hubrich (2005), single country studies, e.g. Reijer and Vlaar (2006), Duarte and Rua (2007), Bruneau et al (2007), Moser, Rumler, and Scharler (2007), more recent studies: Espas and Mayo-Burgos (2013), Bermingham and D'Agostino (2014), Dèes and Güntner (2016).
  - spatial aggregation: Marcellino, Stock and Watson (2003) find that aggregation of country forecasts can improve the accuracy of aggregate forecasts
  - combination of both approaches: Benalal et al. (2004) find that direct approach is better for long run projections, while bottom-up approach is better in the short term
- Density forecast evaluation favouring bottom-up approach:
  - Ravazzolo and Vahey (2014), Tallmand and Zaman (2017): US PCE
  - Mazur (2016): Polish CPI
  - Cobb (2019): France, Germany, UK, forecasts for GDP and CPI are presented.

The models

- 3 large BVAR models based on Giannone et al. (2014):
  - **component model:** components are aggregated based on HICP weights to derive a forecast for HL/HEX inflation
  - **aggregate model for HICP inflation**
  - **aggregate model for HICPX inflation**
- real-time monthly data mimicking the quarterly Eurosystem/ECB staff projection exercises from June 2005 to March 2019 = **56 data vintages**
- EA, and also country results for Big 5 (FR, DE, IT, ES, NL)
- all data starts in January 1997, 100 observations for the oldest vintage, up to 264 obs. for the most recent

Main findings

- **Component vs Aggregate model:**
  - **Point forecasts**: usually no statistically significant difference between aggregate and component model
  - **Density forecasts**: log scores show aggregate model slightly better at most horizons
  - Generally poor predictive densities for all but the shortest horizons in the aggregate model
- **BVAR vs Eurosystem/ECB staff projections:**
  Similar point forecast performance, in the near term Eurosystem is better

Data

Table: Different BVAR specifications

Aggregate Models	HICP Overall index HICP All-items excl. food and energy
Component Model	HICP Unprocessed food HICP Processed food (incl. alcohol and tobacco) HICP Non-energy industrial goods HICP Services
All models	HICP Energy PPI (domestic sales, consumer goods industry) Unit labor costs (whole economy) Non-energy commodity prices: Food (in USD) Non-energy commodity prices (in USD) Nominal effective exchange rate Oil price (in USD) EUR/USD Exchange rate Compensation per employee Real GDP

Forecast evaluation

- **Root mean-squared error:**

$$RMSE_h = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\hat{y}_{r+h} - y_{r+h})^2}$$

- **Log predictive score:**

$$l_h(y_{t+h}) = \frac{1}{R}\sum_{r=1}^{R}\log \hat{p}_r(y_{t+h} \mid y_t)$$

- **Probability Integral Transform (PIT):**

$$u_{t+h} = \int_{-\infty}^{y_{t+h}} \hat{p}(x \mid y_t)dx \equiv P(y_{t+h} \mid y_t)$$

- Diebold-Mariano/Amisano-Giacomini tests of equal forecast accuracy, Harvey, Leybourne, and Newbold (1997) correction
- Battery of calibration tests on PITs

HL: Component model vs aggregate model: point forecasts

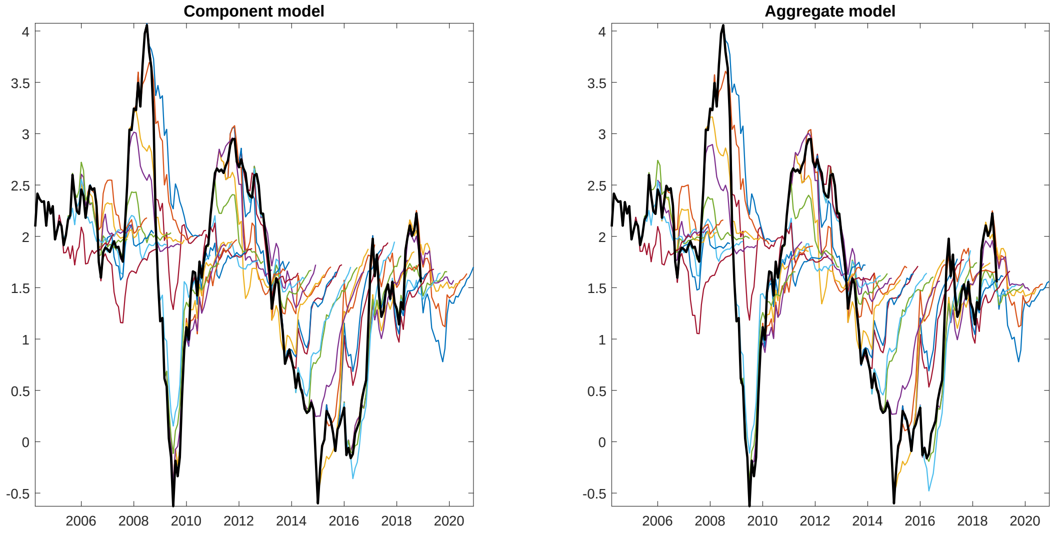


Figure: Conditional forecasts of headline inflation over time

HL: Component vs aggregate model conditional forecasts

Table: Component vs Aggregate model conditional forecasts for headline inflation

Months ahead	RMSE *Component model*	relative RMSE *Component/Aggregate*	relative RMSE *Component/(B)MPE*
1	0.10	0.98	–
3	0.31	0.96	**1.53**
6	0.58	0.97	1.08
12	0.93	0.99	1.15
24	1.06	0.99	1.06
36	1.18	1.02	–

Note: A relative RMSE $<$ 1 indicates that Component model forecasts are more accurate. Bold text denotes statistical significance of the difference at the 5% level, based on Diebold–Mariano tests with Harvey et al. (1997) correction. The comparison with Eurosystem/ECB staff (B)MPE projections (last column) is made on quarterly, rather than monthly, year-on-year inflation rates.

HL: Component model vs aggregate model: density forecasts

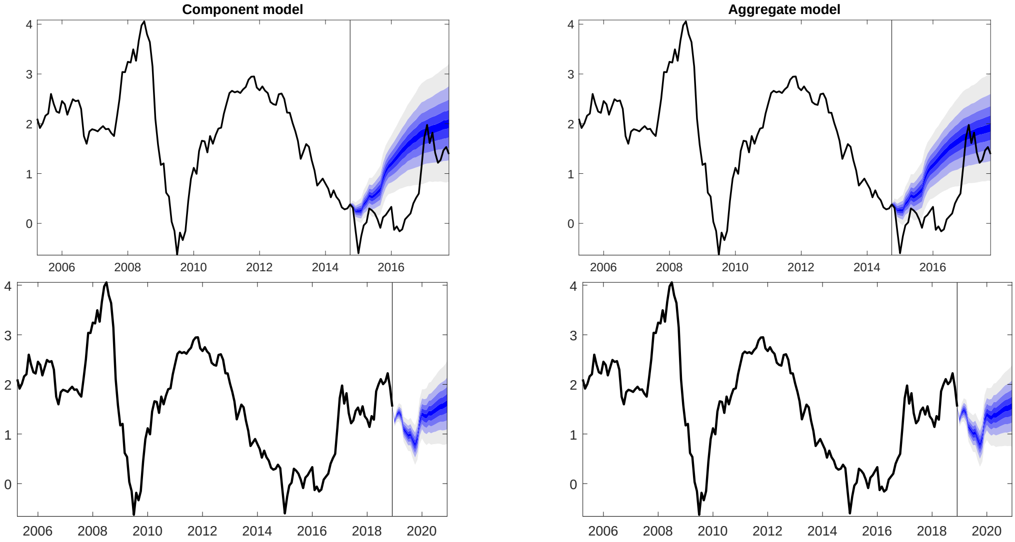
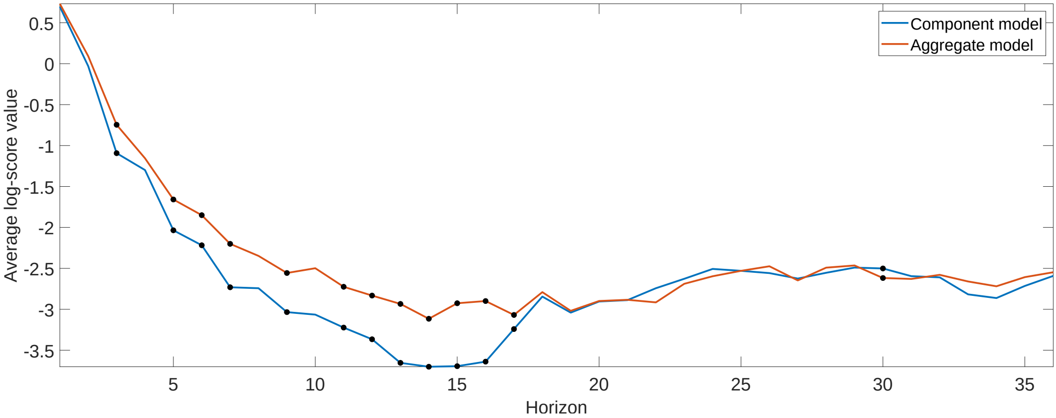


Figure: Fan charts with conditional forecasts of headline inflation (top panel - vintage December 2014, bottom panel - vintage March 2019)

HL: Component model vs aggregate model: Log scores

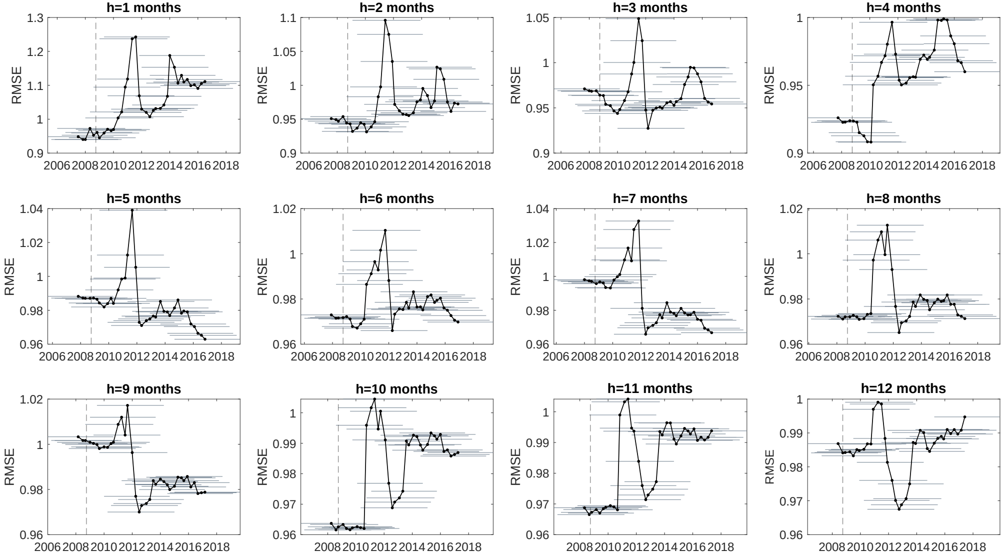
Figure: Average log predictive scores for conditional forecasts - headline inflation



Note: A higher (less negative) log predictive score indicates better forecast accuracy. Black squares mark horizons for which the two scores are statistically different from each other based on Amisano and Giacomini (2007) with Harvey et al. (1997) correction.

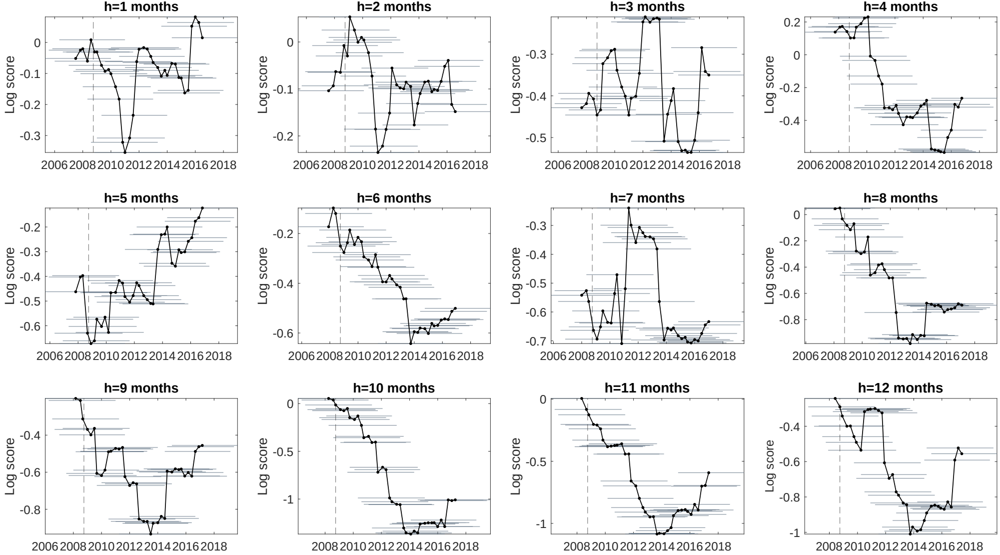
HL: Rolling relative RMSE

Figure: Rolling relative RMSE (component vs aggregate model) by forecast horizon for headline inflation show an improvement of the aggregate model around 2011



HL: Rolling relative log scores

Figure: Rolling relative log scores (component vs aggregate model) by forecast horizon for headline inflation



HL: Calibration tests

Table: Calibration tests on Probability Integral Transforms for headline inflation models

	Aggregate model				**Component model**			
	Uniformity		*Independence*	*Joint H_0*	*Uniformity*		*Independence*	*Joint H_0*
h	*KS*	*AD*	*LB*	*Ber*	*KS*	*AD*	*LB*	*Ber*
1	0.670	0.442	0.964	0.525	**0.039**	**0.009**	0.832	**0.004**
3	**0.003**	<**0.001**	0.316	<**0.001**	**0.009**	<**0.001**	0.458	<**0.001**
6	**0.004**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**
12	**0.003**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**
24	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**
36	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**	<**0.001**

Note: P-values (for horizons $h > 2$, minimum p-values) of the respective test: Kolmogorov-Smirnov (KS), Anderson-Darling (AD), Ljung-Box of the mean (LB), Berkowitz (Ber). Bold text indicates rejection at the 5% significance level (for horizons $h > 2$, using Bonferroni bounds).

HICP excluding food and energy: Main findings

- Point and density forecasts from top-down and bottom-up approaches qualitatively similar
- Differences in point forecast performance typically not statistically significant
- No statistically significant differences from Eurosystem/ECB staff projections
- Top-down model yields better density forecasts across all forecast horizons (small, but statistically significant differences)
- Formal tests on PITs similar to headline (aggregate somewhat better in the near term, otherwise both poorly calibrated)

Countries: Main findings

- Results for big 5 countries are mixed, but with some common threads
- Bottom-up approach usually yields better headline inflation point forecasts in the first months short run, differences then tend to disappear
- Top-down model sometimes delivers better density forecasts
- For HICPX inflation, point forecasts are typically not statistically different; where they are, the top-down model is superior.
- Log scores for HICPX very similar across models, except for Italy, where the aggregate model clearly dominates the other approach for *longer* horizons

Conclusions

- We compare aggregate and component-based conditional inflation forecasts for HICP(X) inflation in the EA and big 5 countries using BVAR models
- EA point forecasts perform similarly, but for density forecasts the top-down approach is better at most horizons
- Point forecasts from our models are as accurate as Eurosystem/ECB staff forecasts, except in the short term, where the latter dominates
- For headline inflation and most forecast horizons, the aggregate model only performs obviously better for a short set of rolling forecast windows centered around the 2011 period; for HICPX, there is no clear change in ranking following the crisis
- For individual countries, the two approaches yield similar results for both inflation measures and both point and density forecasts, with the exception of very short-term point forecasts, where the bottom-up model tends to be better