EVALUATING INFLATION FORECASTS IN THE EURO AREA AND THE ROLE OF THE ECB

Bertrand Candelon, Francesco Roccazzella

REPRINT • 2024 / 12

UCLouvain

LFIN | LiDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

LFIN

Voie du Roman Pays 34, L1.03.01 B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/publications.html

Journal of Forecasting

WILEY

RESEARCH ARTICLE

Evaluating Inflation Forecasts in the Euro Area and the Role of the ECB

Bertrand Candelon[1,2] | Francesco Roccazzella[3]

[1]LFIN/LIDAM, UCLouvain, Ottignies-Louvain-la-Neuve, Belgium | [2]Department of Economics, Maastricht University, Maastricht, Netherlands | [3]IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, Lille, France

Correspondence: Francesco Roccazzella (f.roccazzella@ieseg.fr)

Received: 24 October 2024 | **Accepted:** 29 October 2024

Keywords: ECB | euro area | forecast combinations | forecast evaluation | inflation

ABSTRACT

This paper evaluates the informative value of the ECB inflation forecasts vis-à-vis other institutional and model-based forecasts in the euro area using ex post optimal combinations of forecasts and nonnegative weights. From a methodological perspective, we adapt the corresponding forecast encompassing test to the constrained parameter space, showcasing its superior performance over traditional encompassing tests in both size and power properties. Empirically, the combining weights and the forecast encompassing test reveal that the ECB was the most informative forecaster of euro area inflation over the 2009–2021 period. This changed in 2022: The ECB lost its position as the most informative forecaster, and when using rolling windows to estimate the combining weights using a rolling window, we find an important decline in the ECB's weight over time. This time dependency can be associated with the economic environment and, in particular, the level of uncertainty, the monetary policy, and the macro-financial conditions in which the ECB operates.

Jel classification: E44, E47, E58

1 | Introduction

Forecasting inflation is essential for policymakers to set up their policies but also for investors interested in expected real return, labor market actors during the wage negotiation processes, or households and firms, desiring to evaluate their purchasing power, their saving capacities, or planning their inventory. We can find multiple forecasts building upon time-series models that represent reduced forms of established relationships like the Phillips curve (Stock and Watson 1999) or other forecasts published in surveys by households, firms, professional forecasters, financial market participants, and central bankers. These last ones may combine both quantitative forecasts and expert judgment and provide the best prediction at time t given the available information.

Extensive research has examined the features and the performance of forecasts originating from various sources. For example, focusing on inflation expectations in the United States, Cornand and Hubert (2020) reveal that forecasts from participants in experiments, households, industry and professional forecasters, financial market players, and central bankers exhibit common traits. These include biased forecasts, potential autocorrelation in forecast errors, (in particular over extended forecasting horizons), and the fact that lagged inflation significantly predicts inflation expectations. Another example is the study of Ang, Bekaert, and Wei (2007) that evaluates time series forecasting methods of US inflation vis-à-vis the Survey of Professional Forecasters (SPF) focusing, instead, on forecast combinations and forecast encompassing tests to investigate whether the ex-post optimal[1] combination incorporates more information than the individual forecasts (Clements and Hendry 1998). Their findings indicate that the SPF is the most effective forecasting method compared to time-series approaches and survey information consistently receives the greatest weighting in the forecast combination. Similarly to finance's portfolio theory, the weights reflect forecasts' marginal contribution to the reduction of the forecaster's loss function (Roccazzella, Gambetti, and Vrins 2022), and therefore, when looking at forecast encompassing and forecast combination from a joint perspective, weights can be naturally interpreted as the marginal contribution of each forecaster in term of independent information.

Moving to the euro area (EA), Diebold and Shin (2019) bring evidence that optimal combining weights for the EA inflation projections produced by individual forecasters within the SPF vary over time. Therefore, the forecasts' informational value is equally nonconstant over time. It is yet to be determined whether this finding also applies to institutional forecasters such as the ECB or the IMF and what factors contribute to the dynamics of the informational value of the forecasts over time.

In this paper, we explicitly tackle this point by testing for forecast encompassing and using forecast combinations in the spirit of Bekaert, Hoerova, and Lo Duca (2013) and by investigating the potential source of time variation in the optimal combining weights.

First, we include inflation forecasts from institutions such as the ECB, the European Commission (EC), the International Monetary Fund (IMF), and the Organisation for Economic Co-operation and Development (OECD) in addition to basic time series indicators and other surveys previously employed in the literature. This extension provides new insights into the relative informational advantage of the ECB inflation forecasts compared to the benchmark that the ECB considers when publishing the *European Central Bank staff macroeconomic projection for the euro area*.

Second, we focus on forecast encompassing by estimating the ex-post optimal combination of forecasts that minimizes the combined mean square forecast error while restricting the weights to be nonnegative. From a methodological perspective, we adapt the corresponding test to the constrained parameter space, showing that this outperforms a traditional encompassing test that neglects such constraints in terms of both size and power properties. This novel testing procedure can be extended to frameworks where the number of competing forecasts exceeds the available realizations of the target variable.

Third, in examining the combination weights before 2022, the ECB displays on average the greatest weight. However, no individual inflation forecast, including the ECB, fully encompasses the competitors in a statistically significant way. When the analysis includes 2022, there is a significant shift in results: the ECB loses its status as the most informative forecaster and its weight becomes statistically insignificant at a 5% significance level across various optimal forecast combination strategies. Additionally, when re-estimating the forecast combination using a rolling window approach, it becomes evident that the weight assigned to the ECB has decreased substantially over time.

Fourth, we investigate the sources of this time variability. Naturally, the inflation forecasts generated by the central bank are expected to be the most informative, as they may contain private information on the future monetary policy (Svensson 2005; Faust and Wright 2013). Therefore, studying in what economic states the ECB inflation forecast is more informative can also help policy analysts interested in evaluating the ECB's intent at monitoring inflation.

In this regard, we extend the investigation initiated by Granziera, Jalasjoki, and Paloviita (2021) and Kontogeorgos and Lambrias (2022), which primarily focused on the statistical features of the Eurosystem/ECB staff macroeconomic projections, showing that ECB forecasts are, on average, unbiased and efficient, albeit with strong evidence of state dependence.[2] We show that the informative content of the ECB inflation forecast vis-à-vis other institutional and model-based forecasts is also time-varying and the time-dependency can be associated with the economic environment, and in particular the level of uncertainty, monetary policy, and macro-financial conditions, in which the ECB operates.

The remainder of the paper is organized as follows. Section 2 introduces the optimal combination of forecasts and the forecast encompassing test. Section 3 deals with the analysis of inflation forecasts in the EA whereas Section 4 analyses weights' dynamics. Section 5 concludes.

2 | Optimal Combinations and Forecast Encompassing

In order to evaluate the informational value of inflation forecasts in the EA and test for forecast encompassing, we need to estimate the ex-post optimal combination of forecasts and test the statistical significance of the combining weights. In the following paragraphs, we outline the optimal forecast combination strategy and introduce our forecast encompassing test.

Under a quadratic loss function, it is optimal to minimize the mean square forecast error (Bates and Granger 1969; Granger and Ramanathan 1984). Rather than employing the unconstrained solution, we opt to constrain weights to be nonnegative, as this choice provides three distinct advantages. First, it rules out extreme solutions resulting from estimation error when dealing with the minimum-variance objective function in small samples (Jagannathan and Ma 2003; Conflitti, Mol, and Giannone 2015). Second, it identifies and excludes noise forecasts, i.e., forecasts that individually do not help to predict the target, but whose inclusion in the combination of forecasts is only justified by the objective of minimizing the *in-sample* mean square prediction errors.[3] Third, the optimal weights turn into a special case of the Lasso (Tibshirani 1996), which permits us to extend the test for forecast encompassing to high dimensional problems by encouraging sparse combinations, i.e., combinations with only a few non-zero weights (see for example Brodie et al. 2009).[4]

Formally, given $\boldsymbol{n}$ observations and $\boldsymbol{d}$ forecasts, we denote with $\boldsymbol{y}$ the column vector comprising n observations of the target variable and denote with $\boldsymbol{F}$ and $\boldsymbol{V}$ the $n \times d$ matrices grouping the forecasts and forecast errors, respectively.[5] The optimal nonnegative weights that minimize the mean of the square aggregate forecast error solve

$$\hat{\boldsymbol{\beta}} := \underset{\beta \in \Delta^{d-1}}{\operatorname{argmin}} \frac{1}{n} [\boldsymbol{V}\boldsymbol{\beta}]^T [\boldsymbol{V}\boldsymbol{\beta}] , \qquad (1)$$

with the unit simplex with d vertices defined as

$$\Delta^{d-1} = \left\{ \beta_i \geq 0, \text{ with } i = 1, \ldots, d \text{ and } \sum_{i=1}^{d} \beta_i = 1 \right\} .$$

Using the optimal combining weights estimator derived in (1), we can formulate a forecast encompassing test that explicitly considers the constrained parameter space. Similarly to Harvey and Newbold (2000), we aim to test whether the marginal contribution of forecast i to the aggregate ex-post optimal forecast is statistically significant. Therefore, we state the following hypotheses

$$H_0: \boldsymbol{\theta} = 0 \; vs \; H_1: \boldsymbol{\theta} > 0, \; \boldsymbol{\theta} \in \mathbb{R}^{+(r)} , \tag{2}$$

given the partition $\boldsymbol{\beta} = (\boldsymbol{\theta}, \boldsymbol{\gamma})$, with $\boldsymbol{\theta}$ being the r–dimensional parameter of interest and $\boldsymbol{\gamma}$ denoting the nuisance parameters (with dimension $d - r$). This corresponds to test whether the population ex-post optimal weights lay on the boundary set (under H_0) or in the cone[6] generated by the nonnegativity and the sum to one constraint (under H_1).

Let $\boldsymbol{\epsilon}$ be the n-dimensional vector of forecast error associated with the ex-post optimal combination of forecasts, i.e., $\boldsymbol{\epsilon}: = \boldsymbol{y} - c - \boldsymbol{F}\boldsymbol{\beta}$, with c being the intercept taking into account the presence of bias. Denoting the Gaussian quasi-loglikelihood[7] as

$$\ell(\boldsymbol{\beta}) = -\frac{1}{n} \sum_{i=1}^{n} \log \mathcal{L}_i(\boldsymbol{\beta}) = \frac{n}{2} log(\pi \sigma_{\epsilon}) + \frac{1}{2} \frac{[\boldsymbol{\epsilon}]^T [\boldsymbol{\epsilon}]}{\sigma_{\epsilon}} , \tag{3}$$

the corresponding score statistics is[8]

$$T_S = \left[\nabla \ell \left(\hat{\theta}_{H_0} \right) - \nabla \ell \left(\hat{\theta}_{H_1} \right) \right]^T \widetilde{\mathscr{F}}_{\theta|\gamma}^{-1} \left[\nabla \ell \left(\hat{\theta}_{H_0} \right) - \nabla \ell \left(\hat{\theta}_{H_1} \right) \right] . \tag{4}$$

Under these assumptions, Kudô (1963) proved that the score (and the asymptotically equivalent Wald and likelihood ratio) statistics associated with (2) are distributed as a mixture of chi-squared distributions under the null hypothesis. Therefore, determining the mixing weights becomes essential to compute the cumulative tail probability of the test statistics for a given significance level a. When the constrained space is the nonnegative orthant, Silvapulle and Sen (2004) propose a simple quadratic programming problem to estimate the vector of mixing weights ω[9] and given a significance level a and its corresponding critical value c_a, we compute cumulative tail probability as:

$$\text{Prob}\left[\overline{\chi}^2 (\widetilde{\mathscr{F}}_{\theta|\gamma}, \mathbb{R}^{+r}) > c_a \right] = \sum_{i=0}^{r} \omega_i \text{Prob}\left[\chi^2(i) > c_a \right] ,$$

and we reject H_0 if

$$\text{Prob}\left[T_S > c \right] > a \quad ,$$

where $\chi^2(df)$ denotes a chi-squared distribution with df degrees of freedom. When θ is univariate ($r = 1$), the weights are known in analytical form and the cumulative tail probability reduces to

$$\text{Prob}\left[\overline{\chi}^2 (\widetilde{\mathscr{F}}_{\theta|\gamma}, \mathbb{R}^{+}) > c_a \right] = \frac{1}{2} \text{Prob}\left[\chi^2(1) > c_a \right] . \tag{5}$$

In the remainder of the paper, to ease presentation and simplify notation, we consider the univariate parameter of interest θ ($r = 1$) corresponding to the j^{th} forecast in a low dimensional setting. However, large data sets, heterogeneous modeling techniques, and, different data pre-processing procedures originate a wide array of forecasts, whose size can potentially outclass the number of available realizations of the target variable. Therefore, a "modern" test for forecast encompassing should also deal with potential high dimensional frameworks. This test can accommodate these features using the decorrelated score procedure of Ning and Liu (2017) as described by Yu, Gupta, and Kolar (2019). We refer to Appendix A for the extension of the test in the high dimensional case. In an extensive simulation study in Appendix B, we study the size and size-adjusted power properties of the test, showing how modifying the forecast encompassing test to accommodate the nonnegativity constraint leads to major improvements in both size and size-adjusted power properties compared to disregarding this constraint, also in scenarios involving autocorrelated forecast errors.

3 | Analysis of Inflation Forecasts in the EA

We evaluate nine forecasts for the years ranging from 2009 to 2022.[10] Five are produced by institutions or professional forecasters.[11] We include the ECB forecasts and other institutions or professional forecasters that the ECB considers as a benchmark when publishing Eurosystem staff macroeconomic projections for the EA. Four, instead, are built by simple, yet standard time-series models to forecast inflation traditionally used as benchmarks (Stock and Watson 1999; Bekaert, Hoerova, and Lo Duca 2013). In the spirit of Atkeson and Ohanian (2001), we take a random walk model (*RW*), then we include one autoregressive integrated moving average (*ARIMA*) model and two bi-variate vector-autoregressive (*VAR*) models.

We use the growth rate of the average HICP index in the reference year to measure yearly inflation as the HICP growth rate is the official measure of inflation in the EA and the target of the considered institutional forecasts of inflation.[12] For each quarter, we consider inflation forecasts for the same calendar year because, compared to standard forecasting exercises, institutional forecasters do not publish forecasts with a fixed forecast horizon, e.g., one year ahead, but instead look at calendar years, e.g., the yearly inflation rate in 2018. We avoid longer-term forecasts (1 and 2 calendar years ahead) because there is strong evidence that institutional forecasters fail to systematically update their predictions beyond one calendar year ahead (Andrade and Le Bihan 2013; Easaw and Golinelli 2021).

Institutional forecasts data are retrieved from *European Central Bank staff macroeconomic projection for the euro area*[13] and consist of forecasts produced by the ECB staff, the average forecast published in the SPF, and the ones reported by the EC, the IMF, and the OECD. The ECB, the EC, and the SPF release their inflation forecasts quarterly, while the IMF and OECD publish inflation projections at minimum twice per year in their respective economic outlooks. For each quarter we consider the last available projection of the IMF and the OECD as the respective forecast of inflation. This offers an informative advantage to institutional forecasters that update their expectations more frequently, which can also be reflected in a greater combination weight.[14]

We instead use the EA real-time database (RTDB) to conduct our forecasting exercise with the RW, the ARIMA, and the VARs models, assuming that at any month t we have access to data from the previous months on the variables of interest.[15] The RW and the ARIMA models are estimated via maximum likelihood, instead the two bivariate VAR models are estimated using OLS.[16] The first VAR model captures the joint dynamics of monthly year-on-year (y-o-y) HICP growth rate and industrial production, while the second, focuses on the y-o-y HICP growth and the unemployment rate. The bi-variate VARs align with forecasting strategies based on the generalized Phillips curve, aiming to forecast inflation using real activity measures or unemployment rates (Bekaert, Hoerova, and Lo Duca 2013). We finally combine the monthly y-o-y inflation rates observed in the reference year with iterative forecasts for the subsequent months to form the forecast of the calendar inflation rate. To illustrate, take March 2020 as an example. Utilizing HICP index data preceding March 2020, we project the y-o-y inflation rate for March and the remaining nine months of the year. These predictions, combined with observed y-o-y inflation rates in January and February, are averaged to derive the forecast for the annual inflation rate.

While our model-based forecasts can be updated monthly for the reference calendar year, institutional forecasters release their forecasts at different dates during each quarter. For instance, in Q2, the EC publishes its forecasts in May while the ECB in June. Therefore, estimating the weights at the quarterly frequency by considering the most recent release for each forecaster could give an advantage to institutions that publish their forecasts later in the quarter. However, estimating the weights at monthly frequency might result in strong seasonal patterns: for each forecaster, the weight may display a peak when the corresponding new (updated) forecast is released or a dip when any other forecast is updated. The figures in Appendix D (Figures A4, A5) clearly illustrate this point: an updated forecast is naturally more informative simply because of the publication date. To tackle this problem, we assume that the informational advantage of the monthly update of a forecast later in the quarter is constant and therefore can be averaged out after having performed our analysis for each month in the quarter using the most recent forecasts at the end of the corresponding month. We verify ex post the validity of this assumption by confirming that the peaks and the dips of the weights within each quarter display a constant seasonal pattern as shown in Appendix D. Specifically, we estimate the combining weights (and the corresponding statistics) three times: first, when considering only the most recent forecast at the end of the first month of the quarter; second, when considering only the most recent forecast at the end of the second month of the quarter, and finally when considering only the most recent forecast at the end of the third month of the quarter. The weights (and statistics) used for the analysis will be the simple average of these values.

Table 1 reports the summary statistics of the annual HICP growth rate and its forecasts. The ECB forecasts yield the lowest mean absolute error (MAE) and root mean square error (RMSE) in the considered pool. Following the approach of Romer and Romer (2000), we analyze the mean and standard deviation of the forecast errors to explore potential biases. We find that all

TABLE 1 | Summary statistics of the forecasts and realized forecast errors.

	ARIMA	RW	VAR IP	VAR Un	ECB	OECD	SPF	EC	IMF
RMSE	0.522	1.190	1.590	1.648	0.500	0.990	0.819	0.775	0.933
MAE	0.266	0.670	1.086	1.115	0.228	0.464	0.372	0.364	0.505
(a)	Forecast errors								
Mean	0.01	0.01	0.05	−0.43	0.07	0.11	0.14	0.09	0.14
St. dev.	0.53	1.2	0.58	1.42	0.5	0.99	0.81	0.78	0.93
Med.	−0.03	−0.09	0	−0.05	0.01	0.01	0.02	0.03	0.03
Min.	−1.23	−1.63	−1.02	−6.4	−0.95	−1.31	−0.95	−1.88	−1.58
Max.	2.69	5.73	3.3	2.08	2.92	6.22	5.02	4.52	5.02
Range	3.93	7.36	4.32	8.48	3.87	7.53	5.97	6.4	6.6
(b)	Point forecasts								
Mean	1.75	1.74	1.7	2.18	1.68	1.64	1.61	1.66	1.61
St. dev.	1.73	1.12	1.73	1.89	1.75	1.56	1.55	1.62	1.49
Med.	1.53	1.61	1.49	1.56	1.45	1.5	1.4	1.5	1.45
Min.	0.14	0.29	−0.13	0.26	0	0.1	0.1	−0.1	0.1
Max.	8.08	7.45	8.1	8	8.4	8.3	8.3	8.5	8.3
Range	7.94	7.16	8.23	7.74	8.4	8.2	8.2	8.6	8.2

forecasts are unconditionally unbiased, and looking at Figure 1, we confirm that, on average, the considered forecasts track the dynamics of the yearly inflation rate. However, it is worth underlining two facts.

First, forecasts based on the bi-variate VARs tend to be more volatile, displaying a seasonal component, especially at the beginning of each year. This is particularly evident for the 2018–2021 period where for long forecasting horizons (12–9 steps ahead); that is, for forecasts produced at the beginning of the year (in Q1), VAR IP and VAR UN tend to overestimate the calendar year inflation. This feature disappears as the forecasting horizon shrinks and monthly realized y-o-y inflation data is included in the estimation of the annual HICP growth rate. Second, forecasts tend to either jointly overestimate or jointly underestimate inflation in the EA, implying forecast errors to be cross-sectionally dependent. This is not surprising, especially for institutional forecasters, as it is reasonable to assume that before publishing a new forecast, professional forecasters will compare their results with the ones published by other sources. For example, the ECB acknowledges the use of the SPF and market-based forecasts when forming their expectations. In the next section, we further discuss how we quantify and tackle the cross-sectional dependence problem.

3.1 | Testing and Handling Cross-Sectional Dependence

Strong cross-sectional dependence of the forecast error may signal potentially multi-collinear forecast errors and, consequently, may lead to imprecise estimates of the optimal weights. To quantify the degree of cross-sectional dependence of forecast errors, we implement the *CD*-test (Pesaran 2021) from the panel econometrics literature. *CD*-statistics is defined as

$$CD = \sqrt{\frac{2T}{N(N-1)}} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij}, \tag{6}$$

with $\hat{\rho}_{ij}$ being the sample correlation coefficient between the forecast error of model i and j. In Bailey, Holly, and Pesaran (2016), we can find the order property of the average correlation coefficient:

$$\bar{\rho}_N = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho_{ij} = O(N^{2\alpha-2}), \tag{7}$$

where $\alpha \in [0,1]$. First, as recommended in Elhorst, Gross, and Tereanu (2021), we control that the null of weak cross-sectional dependence ($\alpha < 0.5$) cannot be rejected through the *CD*-test ($\hat{\bar{\rho}}$ and *CD*-statistic's columns in Table 2). We cannot reject weak cross-sectional dependence given the high value of $\hat{\bar{\rho}}$. Second, we assess the degree of cross-sectional correlation through the two-step procedure as in Bailey, Holly, and Pesaran (2016).

In Table 2, we show that the cross-sectional correlation of forecast errors is really strong: the estimated α is larger than 0.75, clearly pointing to the presence of common components driving the dynamics of forecast errors. We can reach similar conclusions by looking at the loadings of the first principal component shown in Table 2: As expected in such a case, they are evenly distributed among the different forecast errors, supporting the evidence of a persistent interdependence trend between inflation forecast errors.

3.2 | Estimating the Optimal Combination Weights

An estimate of the matrix of square forecast errors is necessary to calculate the combining weights. This matrix ($\boldsymbol{V}^T\boldsymbol{V}$) can be

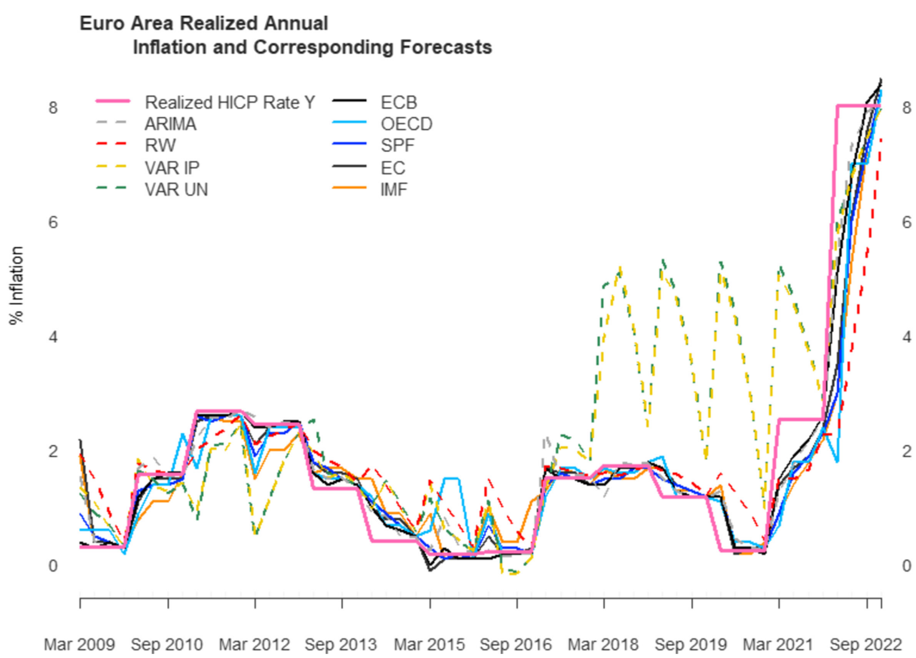

FIGURE 1 | Dynamics of realized and forecasted HICP yearly growth rate.

TABLE 2 | Principal component analysis of realized forecast errors and cross-sectional dependence test.

Forecaster	PC1	PC2	PC3	PC4	PC5
ARIMA R	0.352	−0.142	0.500	0.371	−0.622
RW R	0.357	−0.019	−0.686	0.128	−0.424
VAR IP R	0.188	0.675	0.060	0.025	−0.025
VAR Un R	0.179	0.684	0.058	−0.009	0.039
ECB	0.365	−0.112	0.400	−0.007	0.387
OECD	0.358	−0.095	0.088	−0.850	−0.224
SPF	0.379	−0.108	−0.079	−0.099	0.232
Commission	0.373	−0.117	0.042	0.307	0.243
IMF	0.371	−0.089	−0.311	0.136	0.345
Standard deviation	2.580	1.294	0.483	0.410	0.378
Proportion of variance	0.740	0.186	0.026	0.019	0.016
Cumulative proportion	0.740	0.926	0.952	0.970	0.986
$\hat{\bar{\rho}}$	*CD*-statistic	$\hat{\alpha}$			
0.674	30.275	0.9222			

Note: We report the estimated average sample correlation forecast errors of model i and j, $\hat{\bar{\rho}}$, the cross-sectional test statistics, *CD*-statistics, and $\hat{\alpha}$.

trivially decomposed using the definition of the covariance matrix $\boldsymbol{S}$ as follows:

$$\boldsymbol{S} = \frac{1}{n-d-1}\left[\boldsymbol{V}^T\boldsymbol{V} - \frac{1}{d}\mathbf{1}\mathbf{1}^T\boldsymbol{V}\right]. \tag{8}$$

Notice that in the presence of unbiased forecasts, $\mathbf{1}\mathbf{1}^T\boldsymbol{V}$, that is, the d-dimensional vector of average forecast errors, collapses to a d-dimensional vector of zeros with the corresponding optimal weight minimizing the variance of the aggregate forecast error. To handle the cross-sectional dependence affecting the estimation of the covariance matrix $\boldsymbol{S}$, we start by considering the sample maximum likelihood estimator of the covariance matrix in *Specification 1*. Indeed, as underlined in Jagannathan and Ma (2003) and Conflitti, Mol, and Giannone (2015), restricting weights to lay in the unit simplex helps reduce the risk in the estimated optimal combination of forecasts. However, being a special case of the Lasso (Fan, Zhang, and Yu 2012), this estimator may still fail to correctly select the best subset of forecasts when the forecast errors are highly correlated (Zhao and Yu 2006). For this reason, we consider two additional estimators of the combination weights as robustness checks.

Table 2 suggests the presence of a strong one-factor structure characterizing the dynamics of inflation forecast errors. Following this observation, in *Specification 2*, we consider the principal orthogonal complement thresholding estimator of the covariance matrix (POET) of Fan, Liao, and Mincheva (2013) that assumes that, conditional on pre-specified common factors, the residual terms are weakly correlated.[17]

Finally, we can handle the cross-sectional dependence by imposing some structure on the estimator of the optimal weights. Specifically, we implement a consistent shrinkage estimator of the covariance matrix that combines the sample covariance matrix (which can be easily computed and is asymptotically unbiased but prone to estimation error) with a highly structured estimator (that contains relatively little estimation error but potentially misspecified). Being the loadings of the first principal component roughly the same (Table 2), this factor can further be seen as an equally weighted portfolio of all forecasts (up to a scaling factor). Therefore, it becomes natural to shrink the sample covariance matrix towards a diagonal one (Ledoit and Wolf 2003) that is representative of this case. We estimate the weights adopting such estimator in *Specification 3*.[18]

After having estimated the covariance matrix $\boldsymbol{S}$ using specifications 1, 2, and 3, we can trivially retrieve $\boldsymbol{V}^T\boldsymbol{V}$ using (8) and compute the forecast combination that minimizes the average square forecast error restricting the weights to be nonnegative and to sum to one. The sample used for this analysis spans from the first quarter of 2009 to the fourth quarter of 2022. This last year is unambiguously peculiar. In early 2022, inflation skyrocketed fueled by the important growth rate that followed the end of the COVID crisis and the invasion of Ukraine by the Russian troops, provoking tension in energy markets. 2022 is also the year of the sharp increases in interest rates by the ECB (in September). However, it is still uncertain whether there has been a temporary shift or a structural change in the inflation dynamics in 2022 (Baba et al. 2023) when the inflation rate in Europe reached its highest level, surpassing the expectations of policymakers and exhibiting significant variation across different countries. Therefore, we estimate the weights using a sample that either includes or excludes the 2022 inflation rate. This allows us to assess whether 2022 and the end of the accommodative monetary policy, brought a notable shift in the combining weights.

Table 3 (excluding 2022) and Table 4 (including 2022) report the weights, the T_S score statistics defined in (2) and their corresponding p-values. When estimating the score statistics, we use Newey and West's (1987) HAC estimator of the covariance matrix to cope with the potential autocorrelation of the forecast errors.

Until 2022, it is noticeable that the ECB had (by far) the greatest weight across specifications. This is not surprising as the primary objective of the ECB is to maintain price stability by adjusting monetary policy to make sure that inflation remains low, stable, and predictable. However, the ECB forecast does not encompass the ones of its competitors regardless of the estimation strategy. For example, the ARIMA and the SPF in Spec. 2 are included in the optimal combination of forecasts and their corresponding weights were statistically greater than zero at 5% significance level.

Including 2022 in the estimation sample leads to notably different outcomes. As depicted in Table 4, the ECB loses its position as the most influential forecaster, whereas the ARIMA model displays the highest weight. It is also noteworthy that the other institutional forecasters (IMF, EC, or OECD) have zero or statistically insignificant weights in the ex-post optimal combination of forecasts.

The marked difference in point estimates between Tables 3 and 4 strongly questions the stability of the combining weights over the 2009–2022 period. To challenge this even further, we re-estimate the forecast combination considering a rolling window of 12 quarters.[19]

Figure 2 illustrates the evolution of the average weights across specifications from Q1 2012 to Q4 2022. Meanwhile, Figure 3 zooms in on the ECB, detailing the dynamics of its weight estimated using Spec. 1, 2, and 3. In both instances, the x-axis denotes the reference period; for example, the interval 2015–2016 identifies the weights corresponding to the yearly inflation in 2015. Both figures indicate the fluctuation of weights linked with the ECB and other forecasts over time. We also observe

TABLE 3 | 2009Q1–2021Q4: ex post optimal weights with nonnegativity and sum to one constraints (β), score statistics (T_S), and p-values ($p\overline{\chi}^2$).

	RW	ARIMA	VAR IP	VAR Un	ECB	OECD	SPF	EC	IMF
Spec. 1 – β	0	0.048	0.014	0	0.729	0	0.124	0.085	0
T_S	0	2.691	0.363	0	2.092	0	1.520	1.656	0
$p\overline{\chi}^2$	1	0.050	0.273	1	0.074	1	0.109	0.099	1
Spec. 2 – β	0	0.135	0.064	0.004	0.594	0	0.104	0.080	0.020
T_S	0	4.459	0.445	0.519	3.925	0	3.913	1.900	2.052
$p\overline{\chi}^2$	1	0.017	0.252	0.236	0.024	1	0.024	0.084	0.076
Spec. 3 – β	0	0.097	0.015	0.002	0.690	0	0.108	0.087	0.001
T_S	0	3.992	0.405	0.302	3.381	0	1.505	1.661	1.815
$p\overline{\chi}^2$	1	0.023	0.262	0.291	0.033	1	0.110	0.099	0.089

Note: Spec 1: sample covariance, Spec 2: POET, and Spec 3: L&W shrinkage toward diagonal matrix. Score statistics are computed using Newey and West's (1987) heteroskedasticity and autocorrelation consistent covariance matrix.

TABLE 4 | 2009Q1—2022Q4: ex post optimal weights with nonnegativity and sum to one constraints (β), score statistics (T_S), and p-values ($p\overline{\chi}^2$).

	RW	ARIMA	VAR IP	VAR Un	ECB	OECD	SPF	EC	IMF
Spec. 1 – β	0	0.713	0	0.043	0.243	0	0	0	0
T_S	0	1.650	0	1.350	0.693	0	0	0	0
$p\overline{\chi}^2$	1	0.099	1	0.123	0.203	1	1	1	1
Spec. 2 – β	0	0.539	0.013	0.105	0.289	0	0	0.054	0.0002
T_S	0	2.215	2.267	1.764	2.038	0	0	0.648	0.7103
$p\overline{\chi}^2$	1	0.068	0.066	0.092	0.077	1	1	0.210	0.1997
Spec. 3 – β	0	0.571	0.020	0.036	0.250	0	0.030	0.073	0.020
T_S	0	2.122	2.411	1.910	1.982	0	0.573	0.627	0.703
$p\overline{\chi}^2$	1	0.073	0.060	0.083	0.080	1	0.225	0.214	0.201

Note: Spec 1: sample covariance, Spec 2: POET, and Spec 3: L&W shrinkage toward diagonal matrix. Score statistics are computed using Newey and West's (1987) heteroskedasticity and autocorrelation consistent covariance matrix.

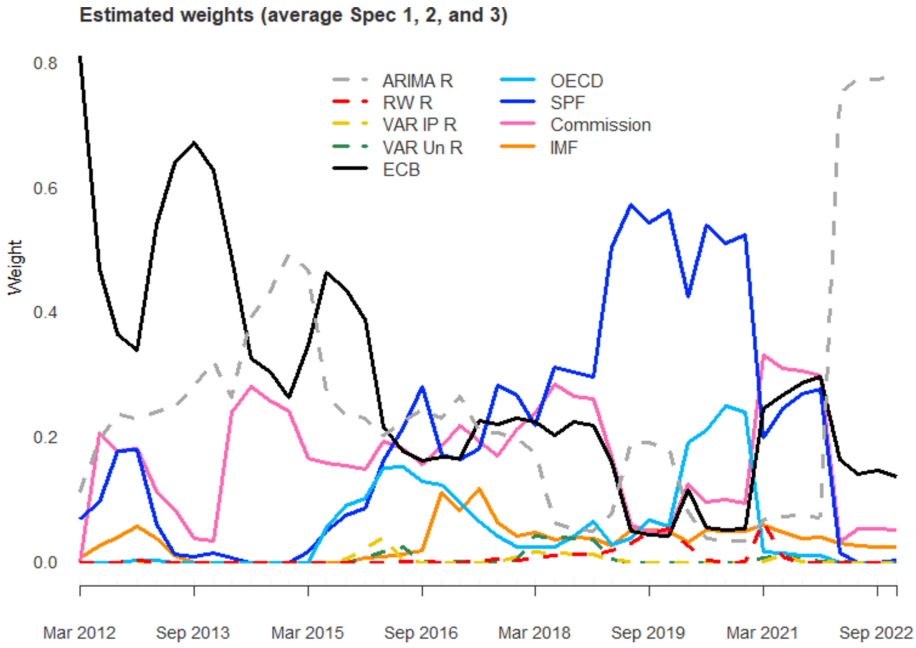

FIGURE 2 | Dynamics of average optimal combination weights across specifications.

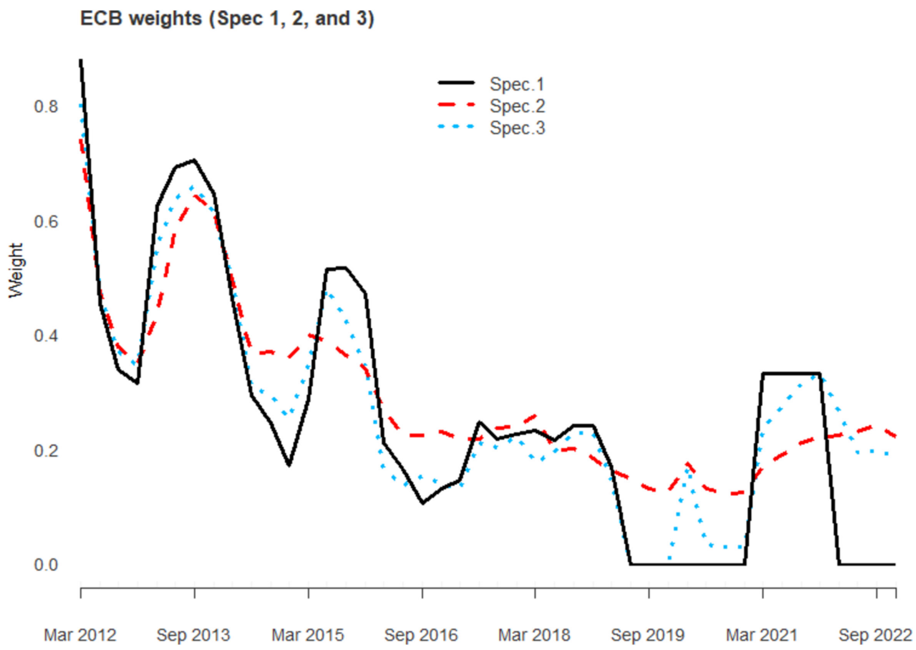

FIGURE 3 | Dynamics of the average ECB weight across specifications.

a downward trend in the weight dynamics associated with the ECB. This observation remains consistent even when considering various subsets of forecasters, such as excluding the bi-variate VARs, the random walk, or some institutional forecasters, as noted in Appendix E. These results underscore the significance of the chosen reference period and how differing conclusions may arise when evaluating unconditional forecasts versus those conditioned on specific economic states. The time-varying nature of the optimal weights can be attributed to various factors, including the diverse underlying processes influencing the HICP in the EA, variations in forecasters' loss functions, and the diversity of information sets available. Hence, it is imperative to investigate whether and how macroeconomic or financial conditions impact the informativeness of ECB forecasts, consequently affecting the ECB's ability to track expected inflation.

4 | The Dynamics of the ECB Weight

Time-varying weights for model-based forecasts could reflect the uncertainty surrounding the forecasting method as, for example, the uncertainty around the choice of the order of the autoregressive and moving average components in ARMA models, the size of the estimation window or the presence of structural breaks in the relation between the time-series in the bivariate *VAR*s. These issues are directly connected to the notion of *heterogeneity* of macroeconomic time series (Clements and Hendry 1998), that is, the nonconstant underlying process generating of the series to forecast. Heterogeneity also matters for institutional forecasters. Being supported by econometric models and expert judgment, the predictive performance of institutional forecasters can be affected by the forecasting techniques used and the turnover of forecasters within the same institution.

The utility function of monetary authority has also important consequences on inflation forecasting. For example, Capistrán (2008) argues that the FED systematic under-prediction and over-predictions of inflation can be explained by the variations of the cost of having inflation above an implicit time-varying target, implying the existence of asymmetric loss functions. Furthermore, while all institutional forecasts are certainly useful for consumers and investors, the ECB macroeconomic projections (of which their inflation forecasts are part) are their primary support for the assessment of economic conjecture and risks to price stability which is the primary objective of the ECB mandate. Therefore, we expect the ECB forecasts to have a direct effect on the monetary policy implemented in the EA. When the expected inflation is low, we expect the monetary authority to try to keep it low to pursue its goal in the medium term. However, policymakers face a dilemma when expected inflation is high: they would like to disinflate, but fear the costs related to a tightening policy (Ball 1992). These costs depend on the state of the economy. In sluggish economic growth, high indebtedness and a vulnerable financial system can strongly amplify the negative effects of the recession that would result from such contractionary policies. In that case, the forecasting community does not know how promptly the monetary authority will react and thus the uncertainty surrounding the inflation dynamics will rise.

Therefore, to examine whether the economic environment in which the ECB operates affects the informativeness of its forecasts, we consider determinants reflecting the uncertainty surrounding the HICP dynamics and macro-financial factors that may affect the ECB loss functions.

4.1 | The Determinants of the Combination Weights

Identifying the exact determinants of the weights is not an easy task. We do not directly observe whether (and when) the HICP time series has experienced structural breaks and how the current (and expected) economic conjecture and risks to price stability have shaped forecasters' loss functions. Nevertheless, we can identify the group of variables that might *indirectly* affect the uncertainty around the expected path of the HICP.

4.1.1 | Disagreement and Inflation Surprise

Following Mankiw, Reis, and Wolfers (2004) and Manzan (2011), we consider the dispersion of (point) forecasts from the SPF to proxy forecasting uncertainty. Patton and Timmermann (2010) empirically show how this measure mostly reflects the heterogeneity in the priors or underlying forecasting models compared to heterogeneity in information signals. Dovern, Fritsche, and Slacalek (2012) confirm that disagreement about inflation forecasts increases with uncertainty about the actual series, but also find that disagreement about prices rises with their level and, in line with the thesis of Rogoff (1985) and Alesina and Gatti (1995), disagreement declines under independent central banks that promote price stability via the introduction of clear mandates in terms of price stability or the adoption of more predictable monetary policy with increased and improved communication. To study whether these aspects influence the dynamics of the ex post optimal weights, we include a measure of the difference between the realized y-o-y monthly HICP growth rate and the 2% target (*Difference from 2%*) and the implied dispersion of (point) forecasts from the SPF *Disagreement SPF*, which we measure as the standard deviation of point forecasts from the ECB SPF.[20] We also control for a broader measure of policy-related economic uncertainty at the European level, employing the news-based Economic policy uncertainty index of Baker, Bloom, and Davis (2016) in our analysis.

Supply shocks, that is, the sudden change of the supply of a product or commodity, resulting in an unforeseen change in price, for example, the surge of gas and oil prices following the Ukrainian war directly affect inflation forecasting performance and, indirectly, the optimal weights. Some forecasters could be better at anticipating or more promptly reacting to recent events because of informative advantages or lower revision costs (Ehrbeck and Waldmann 1996). We proxy the dynamics of inflation surprise with the WTI crude oil price and with the first principal component of realized forecast errors (*Inflation surprise*).

4.1.2 | Financial Conditions and Monetary Policy

The primary objective of the ECB is to maintain price stability by targeting a 2% inflation over the medium term. However, other goals can be pursued "without prejudice to the objective of price stability."[21] These objectives include among others, the stability of the financial system, the risk of financial fragmentation and the risk of euro break up. In this respect, following the outburst of the European sovereign debt crisis, the ECB has taken a range of actions beyond monitoring price stability to address bank funding problems, eliminate excessive risk in sovereign markets, and safeguard monetary policy transmission (Gross and Zahner 2021). These measures were successful at monitoring these risks when the EA was experiencing low (or even negative) inflation (Candelon, Luisi, and Roccazzella 2022; De Grauwe and Ji 2022). However, the Russian invasion of Ukraine, amid an already slowing recovery from the pandemic and inflation pressures, raises new financial stability risks, questions the longer-term impact on economies and the surging commodity prices pose challenging trade-offs for central banks (IMF 2022).

We analyze whether measures of financial stability in the EA affected the past dynamics of the ECB weight. As a proxy of global risk aversion towards bond markets, we consider the long-term corporate *Baa-Aaa* spread (Favero 2013), but also the *global interdependence* and *intra-European fragility* factor of Candelon, Luisi, and Roccazzella (2022). The *global interdependence* factor captures shifts in the level of the entire cross-section of long-term spreads, and it is related to factors that capture devaluation risk (De Santis 2019), flight-to-liquidity mechanisms (Monfort and Renne 2013) and shifts global risk aversion (Favero 2013), further supporting the use of the first *PC* as global trend. The *intra-European fragility* factor, instead, measures how deep the difference between fragile and financially sounder economies in the EA is.

For an exhaustive analysis of financial conditions in the EA, we also include proxies mirroring the monetary policy of the ECB. The shadow rate, as proposed by Wu and Xia (2020), can be considered as a useful proxy for the ECB policy rate when the short-term rate is constrained by the zero lower bound. After September 2022, when the ECB increased its policy rate by 75 basis points, it is adequate to consider the euro short-term rate (ESTER) as a proxy for the ECB policy rate. ESTER is a risk-free overnight interest rate that reflects the wholesale euro unsecured overnight borrowing costs of EA banks. It is computed by the ECB and published on a daily frequency.

As a control, the *APP Factor* that summarizes the growth of Eurosystem holdings corresponding to the ECB asset purchase program (APP) is included. We compute this factor by extracting the first principal component of the y-o-y differences in Eurosystem's holdings purchased under one or more of the APPs operated by the ECB. Specifically, APP includes corporate sector purchase program (CSPP), public sector purchase programme (PSPP), asset-backed securities purchase program (ABSPP), and a third covered bond purchase program (CBPP3).[22]

4.1.3 | Seasonality of Institutional Forecasts

When evaluating institutional forecasts, the literature often overlooks that forecasts always refer to the average of the calendar year.[23] In this framework, seasonal patterns in the dynamics of the estimated weights might emerge, and ignoring them could result in omitted variable bias when studying the relation with their determinants. Therefore, in the spirit of Lovell (1963) and Saikkonen and Lütkepohl (2000), we complete our set of determinants with three seasonal dummy variables corresponding to the second, third, and fourth quarters of the year. This seasonal adjustment implicitly assumes that the intercept of the regression function shifts each quarter and it is convenient to study how the weights are affected when approaching inflation releasing date.

Table 5 summarizes the variables, the corresponding abbreviations, and the sources while we report in Table 6 the correlation matrix between the determinants.

We next introduce the regression models used to study the dynamics of the ECB's weights.

4.2 | The Regression Framework

We study the relationship between the weight of the ECB $\boldsymbol{w}^{\boldsymbol{ECB}}$ and their determinants $\boldsymbol{X}$. Assuming $w_t^{ECB} \sim \mathcal{N}(\boldsymbol{x}_t \beta, \sigma)$, we can estimate the following linear regression model:

$$\boldsymbol{w}^{\boldsymbol{ECB}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \,. \tag{9}$$

Combination weights are, nevertheless, defined in the closed interval [0,1], therefore predictions arising from a linear regression

TABLE 5 | Names and abbreviations of variable employed for the analysis of the determinants of the ex post optimal weight.

Variable	Measure of	Abbreviation	Source
Difference with 2%	Uncertanty	Dif. 2%	Eurostat
Disagreement SPF	Uncertainty	Dis.	ECB
Economic policy uncertainty index	Uncertainty	EPU	Baker, Bloom, and Davis (2016)
Policy rate (shadow + ESTER)	Monetary policy	S	Wu and Xia (2020)
Asset purchasing program factor	Monetary policy	APP	ECB
Cross-sectional forecast error factor	Inflation surprise	Inf. Sur.	Eurostat & ECB
Brent crude oil price	Inflation surprise	Oil	FRED
Spread Aaa-Baa	Financial conditions	Spread	FRED
Global interdependence factor	Financial conditions	Glob.	Candelon, Luisi, and Roccazzella (2022)
Fragility factor	Financial conditions	Frag.	Candelon, Luisi, and Roccazzella (2022)
Seasonal dummy Q2	Seasonality and publ. gap	d_{Q2}	.
Seasonal dummy Q3	Seasonality and publ. gap	d_{Q3}	.
Seasonal dummy Q4	Seasonality and publ. gap	d_{Q4}	.

TABLE 6 | Correlation between the determinants.

	APP	Spread	Oil	Dis.	S	Glob.	Frag.	Dif. 2%	Inf. Sur.
EPU	0.120	0.240	−0.170	0.510	−0.180	−0.240	−0.050	0.6	−0.45
APP		0.230	−0.430	0.010	0.010	0.040	0.16	−0.31	0.07
Spread			−0.260	0.200	0.140	−0.330	0.180	0.01	0.11
Oil				0.150	0.640	−0.560	−0.230	0.47	−0.28
Dis.					0.150	−0.330	−0.100	0.69	−0.53
S						−0.560	−0.130	0.25	0
Glob.							−0.09	−0.47	0.19
Frag.								−0.31	0.27
Dif. 2%									−0.78

framework may lead to negative weights or weights greater than unity. Despite we can adapt the standard linear regression framework to accommodate bounded stochastic processes by transforming the original weights to have values in the real line,[24] this approach has two main shortcomings. First, the regression parameters are interpretable only in terms of the mean of the transformed weights. Second, data generated by bounded stochastic processes are generally skewed and prone to heteroscedasticity.

As a robustness check, we also consider a beta regression model, which is instead tailored to model variables in the interval (0,1).[25] Specifically, we employ the beta regression framework proposed by Ferrari and Cribari-Neto (2004) that assumes (a) independent realizations of $w_i^{ECB} \sim \mathcal{B}(\mu_i, \phi_i)$, (b) a linear regression model for the transform of the mean parameter μ, and (c) a constant precision parameter ϕ. In this framework, we consider a *logit* link function for the mean of w^{ECB}, leading to

$$\mu_i = \frac{e^{\mathbf{x}_i^T \beta}}{1 + e^{\mathbf{x}_i^T \beta}} \ . \tag{10}$$

We refer to Appendix G for more details on the beta regression framework.

To tackle the problems arising from potential heteroscedastic and autocorrelated residuals, we report four-lag Newey-West (NW) standard errors (Newey and West 1987) and a wild bootstrap procedure as in Mammen (1993) in both the linear and beta regression framework.[26] We remark that the linear and beta regression models implicitly accommodate for the estimation uncertainty surrounding the estimated ECB weights. Specifically, if the measurement error in the ECB weights is exogenous to the considered regressors, the regression coefficients remain unbiased and consistent, but the variance of the residual $\boldsymbol{u}$ is expected to increase.

4.3 | Empirical Results

The results on the determinants of the ECB weights obtained using the linear and beta regressions are respectively estimated via OLS and maximum likelihood. We consider four cases. We consider the average ECB weight across specifications, the ECB weight estimated using the sample covariance matrix of prediction errors (Spec. 1), the POET (Spec. 2), and the shrinkage (Spec. 3) estimators.

TABLE 7 | Correlation between the ECB weights estimated using different specifications of the covariance matrix of prediction errors.

	POET	LW
Sample	0.867	0.935
POET	.	0.938

Our study has already revealed in Section 3.2 that the weights' dynamics are similar across different specifications of the covariance matrix of prediction errors, as shown in Figure 3 and by the correlations displayed in Table 7. Thus, it is not surprising that the findings on the determinants of the ECB weight are also consistent across specifications. For the sake of exposition, we focus on the results obtained using the average weight of the ECB across specifications, while we refer to Tables A7, A8, and A9 in the Appendix I for the results built on Specification 1, 2, and 3.

As the sample size is relatively small (only 44 observations), there is a risk of encountering problems such as overfitting and multicollinearity. Therefore, Panel (a) of Table 8 gathers the primary findings of the linear and beta regression model, which concentrate on uncertainty measures, monetary policy rate, and proxies of financial conditions in the EA. In Panel (b), additional analyses including as controls the asset purchasing program factor (APP), *Oil*, *Spread*, and seasonal dummies are reported.

The table displays the estimated coefficients in the linear (LM) and beta regression model (Beta), respectively. The corresponding Newey-West standard errors are reported in *SE*, while the columns NW and B report whether the coefficient of interest is statistically different from 0 at 1% (***), 5% (**), and

TABLE 8 | Determinants of the weights constructed using the average (Spec. 1, 2, and 3) weights.

(a)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.703	0.057	***	***	1.055	0.328	***	***
Dif. 2%	−0.051	0.009	***	***	−0.233	0.044	***	***
EPU	−0.001	0.000	***	***	−0.004	0.001	***	***
S	0.055	0.006	***	***	0.279	0.035	***	***
Inf. Sur.	−0.033	0.006	***	***	−0.176	0.033	***	***
Glob.	−0.009	0.005	*	**	−0.022	0.021		
Frag.	−0.031	0.012	**	***	−0.149	0.053	***	***
R^2	0.855				0.7958			

(b)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.498	0.317		*	1.103	0.581	*	*
Dif. 2%	−0.051	0.016	***	***	−0.229	0.093	**	***
Dis.	−0.006	0.029			−0.116	0.210		
EPU	−0.001	0.000	***	***	−0.004	0.001	***	***
S	0.054	0.006	***	***	0.283	0.045	***	***
APP	−0.001	0.011			−0.023	0.067		
Inf. Sur.	−0.033	0.009	***	***	−0.175	0.053	***	***
Glob.	−0.004	0.008			−0.008	0.033		
Frag.	−0.032	0.009	***	***	−0.147	0.046	***	***
Oil	0.038	0.061			0.130	0.612		
Spread	0.069	0.067			0.286	0.339		
d_2	−0.063	0.027	**	**	−0.331	0.150	**	**
d_3	−0.039	0.028			−0.223	0.155		
d_4	−0.056	0.041		*	−0.314	0.236		*
R^2	0.850				0.7780			

Note: The first column displays the determinants; b_0 is the intercept. LM and Beta report the estimated coefficients using the linear regression model and the beta regression. We report the adjusted R^2 for the LM and adjusted pseudo adjusted R^2 for the Beta. We display Newey-West standard errors in column SE. Columns NW and B report whether the coefficient of interest is statistically significant at 1%, 5%, and 10% levels according to the Newey-West standard errors and the Mammen wild bootstrap quantiles, respectively. We count 44 observations.
*Statistically significant at 10% level.
**Statistically significant at 5% level.
***Statistically significant at 1% level.

10% (*) levels according to the Newey-West corrected standard errors and the Mammen wild bootstrapped quantiles, respectively.

Starting from Panel (a), we find that the measures of uncertainty (*Dis.* and *EPU*), how actual inflation differs from the 2% target (*Dif. 2%*), and inflation surprise (*Infl. Sur.*) are negatively associated to the weight of the ECB. The relationship between the monetary policy rate (*S*) and the informativeness of the ECB is intriguing. It shows that when the policy rate increases, so does the weight, indicating that a more stringent monetary policy on average correlates with a higher informational value in the ECB inflation forecast. Regarding financial conditions in the EA, we find that the fragmentation factor (*Frag.*)—which measures the depth of the difference between fragile and financially sounder economies—is on average negatively associated with the ECB weight.

In Panel (a), looking at the adjusted R^2, we notice that the regression models explain approximately 80% of the weight variability and, for all except *Glob.*, the slope coefficients are statistically significant at least at the 10% level. In Panel (b), we add the controls *Dis.*, *Oil*, *Spread*, the *APP* factor, and the seasonal dummies. However, it is important to note that these controls do not enhance the explanatory power of the regressions, as the adjusted R^2 value decreases in both LM and Beta models, and the slope coefficients associated with the control variables are mostly insignificant. Therefore, when we discuss the results, we will concentrate on the restricted model that excludes the controls.

4.4 | Discussion: Forecast Informativeness and Monetary Policy

In our framework, weights reflect how informative forecasts are: an upward (downward) dynamics of the weight relative to the competitors signals that the considered forecaster has become relatively more (less) informative when producing inflation forecasts. Being more informative, that is, featuring a greater weight, implies that the public should regard that forecaster as more relevant when forecasting the variable of interest. In this perspective, low weights relative to their historical trend and the other forecasters may be interpreted as warning flags, signaling that the informativeness of the ECB inflation forecasts is at risk.

We found that ECB weight declined over time and that it correlates with economic conditions in the EA. In particular, high levels of uncertainty, loose monetary policy, and threats to financial stability were negatively associated with the weight of the ECB.

From the ECB's point of view, this can be a problem. Having informative forecasts (or forecasts that are believed to be informative) is a crucial strategic tool for conducting monetary policy. First, since realized inflation typically responds to changes in monetary policy only with a substantial lag (from one to two years or more), having reliable information about the expected path of the price level offers a substantial timing advantage when designing monetary policy. Second, Svensson (1997) showed that inflation-targeting also implies expected inflation-targeting: the central bank's inflation forecast becomes an explicit intermediate target. This should simplify both the implementation and monitoring of monetary policy, helping the inflation-targeting central bank and the public to tell whether the monetary policy is fulfilling its mandate to maintain price stability (Bernanke and Mishkin 1997). However, this could equally have serious adverse consequences for the central bank's accountability and credibility[27] if the central bank's forecasts were believed to be inconsistent with the expected dynamics of inflation. Therefore, while the decreased weights compared to their historical patterns could be seen as cautionary signals, the correlations between the ECB weight, macro-financial conditions, and monetary policy in the EA could also imply that the tightening of monetary policy in the second part of 2022 (partially captured in the analyzed data) might predict an increase in the informativeness of ECB forecasts. The availability of an extended dataset in the coming years will be crucial in evaluating whether the ECB will regain its position as the most informative inflation forecaster in the EA.

5 | Conclusion

In this paper, we evaluate the relative informational advantage of ECB inflation forecasts compared to those that the ECB includes as a benchmark when publishing the *European Central Bank staff macroeconomic projection for the euro area* like the EC, the IMF, the OECD, in addition to the SPF, and simple benchmark time-series models like random walk, ARIMA and bi-variate VARs.

With this goal, we estimate ex-post optimal combinations of forecasts while restricting the weights to be nonnegative and adapting the corresponding encompassing test to this constrained parameter space. This outperforms a traditional encompassing test that neglects this constraint in terms of size and power properties, while also accommodating frameworks where the number of competing forecasts exceeds the available realizations of the target variable.

After examining the combination weights before 2022, we find that the ECB displays on average the greatest weight, and no individual forecasts of inflation (ECB included) encompass the competing forecasts in the EA. However, results markedly change when including 2022 in the sample: The ECB loses its position as the most informative forecaster. Moreover, when estimating the forecast combinations using rolling windows, we observe an important decline in the ECB's weight over time. We investigate the sources of this time variability and find that macro-financial conditions and monetary policy in the EA correlate with this pattern. Greater uncertainty surrounding inflation and the difference between the current and the 2% target are associated with less informative ECB forecasts. A tighter monetary policy is instead associated with more informative ECB inflation forecasts.

Acknowledgments

The authors are grateful to David Ardia, Francesca Monti, Roberto Renò, Walter Distaso, Christian Brownlees, Frédéric Vrins, Athanasios Triantafyllou, and Raf Wouters, the participants in the 2023 Belgian Financial Research Forum, in the 2024 International Conference on Macroeconomic Analysis and International Finance, the 2024 International Conference of the French Finance Association (AFFI), in the UCLouvain CORE Brown Bag, the research seminars at the University of Alicante, at THEMA - CY Cergy Paris Université, at the Paris School of Economics, at the IESEG School of Management, and the National Bank of Belgium. Bertrand Candelon has conducted as part of the research program entitled "Financial and Extra-financial Risks Modeling" under the aegis of the Europlace Institute of Finance, a joint initiative with Insti7. The usual disclaimer applies. Earlier versions of this article were circulated with the title *Should we care about ECB inflation expectations?*.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] Optimal in the sense of minimizing the mean square error of the combined forecast.

[2] Granziera, Jalasjoki, and Paloviita (2021) and Kontogeorgos and Lambrias (2022) show that the ECB tends to overpredict (underpredict) inflation at intermediate forecast horizons when inflation is below (above) target, with this bias being more pronounced when inflation is above the target.

[3] It is well known that the inclusion of such forecasts can negatively impact the stability of the weights (Bunn 1981).

[4] For example, Fan, Zhang, and Yu (2012) employed this correspondence to estimate approximate weights for the minimum variance portfolio under gross-exposure constraints when the number of securities under consideration is (potentially) larger than the sample size.

[5] In this paper, we denote matrices with capital bold letters, vectors with lowercase bold letters, and scalars with lowercase letters.

[6] A subset C of $\mathbb{R}^d$ is called a cone if $\forall x \in C, tx \in C$ for every positive t.

[7] We discuss the empirical distribution of ϵ in our application in Appendix C.

[8] We let $\nabla \ell(\boldsymbol{\beta}) = \nabla \ell(\theta, \boldsymbol{\gamma})$ be the gradient of the negative log likelihood, $\nabla^2 \ell(\boldsymbol{\beta})$ be the sample Hessian matrix and $\mathscr{F}(\beta) = 34\mathbb{E}[\nabla^2 \ell(\boldsymbol{\beta})]$ be the population Fisher information matrix, while we denote with $\nabla_\theta \ell(\boldsymbol{\beta})$, $\nabla_\gamma \ell(\boldsymbol{\beta})$, $\mathscr{F}(\beta)_{\theta\theta}$, $\mathscr{F}(\beta)_{\theta\gamma}$, $\mathscr{F}(\beta)_{\gamma\theta}$, $\mathscr{F}(\beta)_{\gamma\gamma}$ the correspondent partitions. We also denote with $\hat{\boldsymbol{\beta}} = (\hat{\theta}, \hat{\gamma})$ the estimated ex-post optimal weights.

[9] See proposition 3.6.1 of Silvapulle and Sen (2004).

[10] Considering nine forecasts offers the benefit of analyzing potential time-varying combining weights using a relatively short (twelve-quarters) rolling window while keeping two degrees of freedom to estimate optimal combining weights.

[11] Financial instruments such as inflation-linked swaps (ILS) and inflation-linked bonds (ILB), such as the Italian BTPi indexed Treasury bonds, can also be used as a market-based source of inflation forecasts. In the EA, the Harmonised Index of Consumer Prices excluding tobacco is the relevant index to which both ILS and ILB are anchored. This is a shortcoming of using ILS and ILB-implied inflation forecasts, as they forecast a different measure of inflation compared to the ECB, the SPF, the EC, the IMF, and the OECD that, instead, consider the HICP index. Moreover, ILB and ILS yields embody not only the inflation expectation, but also a risk premium related to inflation uncertainty, counterparty, and (in the case of ILB) credit risk, which can vary over time. ILB and ILS inflation-based forecasts cannot be directly employed without disentangling all these components. For these reasons, this study cannot employ ILS and ILB implied inflation forecasts.

[12] We obtain comparable results when considering the December-on-December change of the realized value of HICP to measure yearly inflation as in Patton and Timmermann (2010). As standard in the forecast encompassing literature, we assume that the target variable is observed without error. However, we remark that in the linear regression model, on which the optimal forecast combination is based, it is possible to accommodate for measurement error in the variable to be forecasted if this is exogenous to the considered forecasts. In this case, the estimator of the weights remains unbiased and consistent, but the variance of the residual ϵ is expected to increase.

[13] Available at https://www.ecb.europa.eu/pub/projections/html/all-releases.en.html

[14] This may explain why OECD and IMF forecasts display worse performance than the other institutional forecasters (see St. Dev. of the forecast errors in Table 1). Nevertheless, our results are robust to the exclusion of these two forecasters and when considering other subsets of institutional and model-based forecasts (see Appendix E, Figure A6).

[15] Institutional forecasters are usually published at the end of the reference month and, therefore, when considering VARs, ARIMA, and RW, we construct forecasts at the end of the reference month with only data from the prior months. This assumption is fair given the monthly time series we are considering. For example, the EA HICP's first ‘flash estimate’ is published around the end of the reference month. Similarly, industrial production and unemployment are usually published 45 and 30 days after the end of the reference month, respectively.

[16] Time-series models are estimated using a rolling window of 48 months. We have also considered rolling windows of 24, 60, and 72 months and expanding windows. We chose the window of 48 because it is the one that offers the best performance in terms of RMSE across specifications. For each of the considered windows, we control for the stationarity of the y-o-y HICP growth rate and select the appropriate differencing order by performing the augmented Dickey-Fuller test. Similarly, we select the appropriate order of the ARIMA and the order of the VAR using the Akaike information criterion (AIC).

[17] We consider only one common factor proxied by the first principal component.

[18] Following Ledoit and Wolf (2003), the weight between the sample and the target matrix is computed by minimizing the Frobenius norm of the difference between the shrinkage estimator and the asymptotic covariance matrix.

[19] As a robustness check, we have considered windows of 16 and 20 quarters. Results are comparable, and we report them in Appendix F (Figures A7, A8).

[20] Results are comparable when we considered alternative measures of dispersion for our measure. Specifically, we also considered interquartile or the 95th–5th range for the *Disagreement SPF*.

[21] See the ECB website for more details: https://www.ecb.europa.eu/mopo/intro/html/index.en.html.

[22] For data and more information, we refer to https://www.ecb.europa.eu/mopo/implement/app/html/index.en.html.

[23] In the first quarter of the year, the current-year forecast is four quarters ahead, in the remaining quarters of the year, the horizon becomes three, two, and one quarter(s) ahead, respectively.

[24] For example via standard *logit* or Box's transformations.

[25] Following Smithson and Verkuilen (2006), we consider the transformation $\frac{w^{ECB}(n-1)+0.5}{n}$ where n is the sample size to adapt the beta regression framework to dependent variables assuming also the extremes 0 and 1.

[26] We refer to Appendix H for a detailed description of the bootstrap algorithm.

[27] The term “credibility” refers to the degree of confidence that the public has in the central bank's determination and ability to meet its announced objectives.

[28] Remark that $\boldsymbol{W}^* = \mathscr{F}^{*-1}_{\gamma\gamma} \mathscr{F}^*_{\gamma\theta}$.

[29] We refer to Appendices C and D of Ning and Liu (2017) for the proofs that Assumptions 1–5 hold in a general linear regression model, while we refer to the appendix of Fan et al. (2012) for the proofs related to the nonnegative least squares estimator under sum to one constraint.

[30] The analysis of the empirical size and size-adjusted power in schemes (a) and (b) are comparable. The parameters $\sigma_z = 0.5$ and $\sigma_u = 0.25$ are chosen to obtain an average signal-to-noise ratio ranging from 1.60 (when $\rho = 0.75$) to 1.8 (when $\rho = 0$) across the simulation study for a sample of 1000 observation. Similarly, we report results for $d = 5$, but other tables with additional checks are available upon request.

References

Alesina, A., and R. Gatti. 1995. “Independent Central Banks: Low Inflation at No Cost?.” *American Economic Review* 85, no. 2: 196–200. http://www.jstor.org/stable/2117918.

Andrade, P., and H. Le Bihan. 2013. “Inattentive Professional Forecasters.” *Journal of Monetary Economics* 60, no. 8: 967–982.

Ang, A., G. Bekaert, and M. Wei. 2007. “Do macro Variables, Asset Markets, or Surveys Forecast Inflation Better?.” *Journal of Monetary Economics* 54, no. 4: 1163–1212. https://www.sciencedirect.com/science/article/pii/S0304393206002303.

Atkeson, A., and L. E. Ohanian. 2001. “Are Phillips Curves Useful for Forecasting Inflation?.” *Quarterly Review* 25, no. 1. https://doi.org/10.21034/qr.2511.

Baba, C., R. Duval, T. Lan, and P. Topalova. 2023. "The 2020–2022 Inflation Surge Across Europe: A Phillips-Curve-Based Dissection." (*IMF Working Papers 030*): International Monetary Fund.

Bailey, N., S. Holly, and M. H. Pesaran. 2016. "A Two-Stage Approach to Spatio-Temporal Analysis With Strong and Weak Cross-Sectional Dependence." *Journal of Applied Econometrics* 31, no. 1: 249–280.

Baker, S. R., N. Bloom, and S. J. Davis. 2016. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics* 131, no. 4: 1593–1636.

Ball, L. 1992. "Why Does High Inflation Raise Inflation Uncertainty?." *Journal of Monetary Economics* 29, no. 3: 371–388.

Bates, J. M., and C. W. J. Granger. 1969. "The Combination of Forecasts." *Journal of the Operational Research Society* 20, no. 4: 451–468.

Bekaert, G., M. Hoerova, and M. Lo Duca. 2013. "Risk, Uncertainty and Monetary Policy." *Journal of Monetary Economics* 60, no. 7: 771–788.

Bernanke, B. S., and F. S. Mishkin. 1997. "Inflation Targeting: A New Framework for Monetary Policy?." *Journal of Economic Perspectives* 11, no. 2: 97–116. https://www.aeaweb.org/articles?id=10.1257/jep.11.2.97.

Brodie, J., I. Daubechies, C. De Mol, D. Giannone, and I. Loris. 2009. "Sparse and Stable Markowitz Portfolios." *Proceedings of the National Academy of Sciences* 106, no. 30: 12267–12272. https://www.pnas.org/content/106/30/12267.

Bunn, D. W. 1981. "Two Methodologies for the Linear Combination of Forecasts." *Journal of the Operational Research Society* 32, no. 3: 213–222.

Busetti, F., and J. Marcucci. 2013. "Comparing Forecast Accuracy: A Monte Carlo Investigation." *International Journal of Forecasting* 29, no. 4: 13–27.

Candelon, B., A. Luisi, and F. Roccazzella. 2022. "Fragmentation in the European Monetary Union: Is It Really Over?." *Journal of International Money and Finance* 122: 102545. https://www.sciencedirect.com/science/article/pii/S0261560621001960.

Capistrán, C. 2008. "Bias in Federal Reserve Inflation Forecasts: Is the Federal Reserve Irrational or Just Cautious?." *Journal of Monetary Economics* 55, no. 8: 1415–1427. https://www.sciencedirect.com/science/article/pii/S0304393208001499.

Clements, M., and D. Hendry. 1998. *Forecasting Economic Time Series*. Cambridge University Press. https://EconPapers.repec.org/RePEc:cup:cbooks:9780521634809.

Conflitti, C., C. D. Mol, and D. Giannone. 2015. "Optimal Combination of Survey Forecasts." *International Journal of Forecasting* 31, no. 4: 1096–1103.

Cornand, C., and P. Hubert. 2020. "On the External Validity of Experimental Inflation Forecasts: A Comparison With Five Categories of Field Expectations." *Journal of Economic Dynamics and Control* 110: 103746. https://www.sciencedirect.com/science/article/pii/S0165188919301459.

De Grauwe, P., and Y. Ji. 2022. "The Fragility of the Eurozone: Has It Disappeared?." *Journal of International Money and Finance* 120: 102546. https://www.sciencedirect.com/science/article/pii/S0261560621001972.

De Santis, R. 2019. "Redenomination Risk." *Journal of Money, Credit and Banking* 51, no. 8: 2173–2206. https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12582.

Diebold, F. X., and M. Shin. 2019. "Machine Learning for Regularized Survey Forecast Combination: Partially-Egalitarian LASSO and Its Derivatives." *International Journal of Forecasting* 35, no. 4: 1679–1691.

Dovern, J., U. Fritsche, and J. Slacalek. 2012. "Disagreement Among Forecasters in G7 Countries." *Review of Economics and Statistics* 94, no. 4: 1081–1096. https://doi.org/10.1162/REST_a_00207.

Easaw, J., and R. Golinelli. 2021. *Professionals Inflation Forecasts: The Two Dimensions of Forecaster Inattentiveness*. Oxford Economic Papers. https://doi.org/10.1093/oep/gpab012.

Ehrbeck, T., and R. Waldmann. 1996. "Why Are Professional Forecasters Biased? Agency Versus Behavioral Explanations." *Quarterly Journal of Economics* 111, no. 1: 21–40. http://www.jstor.org/stable/2946656.

Elhorst, J. P., M. Gross, and E. Tereanu. 2021. "Cross-Sectional Dependence and Spillovers in Space and Time: Where Spatial Econometrics and Global Var Models Meet." *Journal of Economic Surveys* 35, no. 1: 192–226.

Fan, J., Y. Liao, and M. Mincheva. 2013. "Large Covariance Estimation by Thresholding Principal Orthogonal Complements." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 75, no. 4: 603–680. http://www.jstor.org/stable/24772450.

Fan, J., J. Zhang, and K. Yu. 2012. "Vast Portfolio Selection With Gross-Exposure Constraints." *Journal of the American Statistical Association* 107, no. 498: 592–606.

Faust, J., and J. H. Wright. 2013. "Chapter 1—Forecasting Inflation." In *Handbook of Economic Forecasting*, edited by G. Elliott, and A. Timmermann, Handbook of Economic Forecasting, Vol. 2, 2–56. Elsevier.

Favero, C. A. 2013. "Modelling and Forecasting Government Bond Spreads in the Euro Area: A GVAR Model." *Journal of Econometrics* 177, no. 2: 343–356.

Ferrari, S., and F. Cribari-Neto. 2004. "Beta Regression for Modelling Rates and Proportions." *Journal of Applied Statistics* 31, no. 7: 799–815.

Fu, W., and K. Knight. 2000. "Asymptotics for Lasso-Type Estimators." *Annals of Statistics* 28, no. 5: 1356–1378. https://doi.org/10.1214/aos/1015957397.

Gambetti, P., G. Gauthier, and F. Vrins. 2019. "Recovery Rates: Uncertainty Certainly Matters." *Journal of Banking & Finance* 106: 371–383.

Granger, C. W. J., and R. Ramanathan. 1984. "Improved Methods of Combining Forecasts." *Journal of Forecasting* 3, no. 2: 197–204.

Granziera, E., P. Jalasjoki, and M. Paloviita. 2021. "The Bias and Efficiency of the ECB Inflation Projections: A State Dependent Analysis." (*No. 7*): Bank of Finland Research Discussion Paper. https://doi.org/10.2139/ssrn.3837123.

Gross, J., and J. Zahner. 2021. "What Is on the ECB's Mind? Monetary Policy Before and After the Global Financial Crisis." *Journal of Macroeconomics* 68: 103292. https://www.sciencedirect.com/science/article/pii/S0164070421000057.

Harvey, D., and P. Newbold. 2000. "Tests for Multiple Forecast Encompassing." *Journal of Applied Econometrics* 15, no. 5: 471–482. http://www.jstor.org/stable/2678604.

Hyndman, R. J., and Y. Fan. 1996. "Sample Quantiles in Statistical Packages." *American Statistician* 50, no. 4: 361–365. http://www.jstor.org/stable/2684934.

IMF. 2022. *Global Financial Stability Report: Financial Stability Risks Have Risen as War Tests the Resilience of the Financial System Through Various Channels*. IMF.

Jagannathan, R., and T. Ma. 2003. "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps." *Journal of Finance* 58, no. 4: 1651–1683.

Kontogeorgos, G., and K. Lambrias. 2022. "Evaluating the Eurosystem/ECB Staff Macroeconomic Projections: The First 20 Years." *Journal of Forecasting* 41, no. 2: 213–229. https://onlinelibrary.wiley.com/doi/abs/10.1002/for.2804.

Kudô, A. 1963. "A Multivariate Analogue of the One-Sided Test." *Biometrika* 50, no. 3-4: 403–418.

Ledoit, O., and M. Wolf. 2003. "Improved Estimation of the Covariance Matrix of Stock Returns With an Application to Portfolio Selection." *Journal of Empirical Finance* 10, no. 5: 603–621.

Lee, J. D., D. L. Sun, Y. Sun, and J. E. Taylor. 2016. "Exact Post-Selection Inference, With Application to the Lasso." *Annals of Statistics* 44, no. 3: 907–927.

Lovell, M. C. 1963. "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis." *Journal of the American Statistical Association* 58, no. 304: 993–1010. https://www.tandfonline.com/doi/abs/10.1080/01621459.1963.10480682.

Mammen, E. 1993. "Bootstrap and Wild Bootstrap for High Dimensional Linear Models." *Annals of Statistics* 21, no. 1: 255–285.

Mankiw, N. G., R. Reis, and J. Wolfers. 2004. "Disagreement About Inflation Expectations." In *NBER Macroeconomics Annual 2003*, Vol. 18, 209–270. National Bureau of Economic Research, Inc. https://EconPapers.repec.org/RePEc:nbr:nberch:11444.

Manzan, S. 2011. "Differential Interpretation in the Survey of Professional Forecasters." *Journal of Money, Credit and Banking* 43, no. 5: 993–1017.

Monfort, A., and J.-P. Renne. 2013. "Decomposing Euro-Area Sovereign Spreads: Credit and Liquidity risks." *Review of Finance* 18, no. 6: 2103–2151.

Newey, W. K., and K. D. West. 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix." *Econometrica* 55, no. 3: 703. https://doi.org/10.2307/1913610.

Ning, Y., and H. Liu. 2017. "A General Theory of Hypothesis Tests and Confidence Regions for Sparse High Dimensional Models." *Annals of Statistics* 45, no. 1: 158–195.

Patton, A. J., and A. Timmermann. 2010. "Why Do Forecasters Disagree? Lessons From the Term Structure of Cross-Sectional Dispersion." *Journal of Monetary Economics* 57, no. 7: 803–820. https://www.sciencedirect.com/science/article/pii/S0304393210000899.

Pesaran, M. H. 2021. "General Diagnostic Tests for Cross-Sectional Dependence in Panels." *Empirical Economics* 60: 13–50.

Roccazzella, F., P. Gambetti, and F. Vrins. 2022. "Optimal and Robust Combination of Forecasts via Constrained Optimization and Shrinkage." *International Journal of Forecasting* 38, no. 1: 97–116.

Rogoff, K. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100: 1169–1189.

Romer, C. D., and D. H. Romer. 2000. "Federal Reserve Information and the Behavior of Interest Rates." *American Economic Review* 90, no. 3: 429–457. https://www.aeaweb.org/articles?id=10.1257/aer.90.3.429.

Saikkonen, P., and H. Lütkepohl. 2000. "Testing for the Cointegrating Rank of a VAR Process With Structural Shifts." *Journal of Business & Economic Statistics* 18, no. 4: 451.

Silvapulle, M., and P. Sen. 2004. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Wiley.

Smithson, M., and J. Verkuilen. 2006. "A Better Lemon Squeezer? Maximum-Likelihood Regression With Beta-Distributed Dependent Variables." *Psychological Methods* 11: 54–71.

Stock, J. H., and M. W. Watson. 1999. "Forecasting Inflation." *Journal of Monetary Economics* 44, no. 2: 293–335. https://www.sciencedirect.com/science/article/pii/S0304393299000276.

Svensson, L. E. O. 1997. "Inflation Forecast Targeting: Implementing and Monitoring Inflation Targets." *European Economic Review* 41, no. 6: 1111–1146. https://www.sciencedirect.com/science/article/pii/S0014292196000554.

Svensson, L. E. O. 2005. "Monetary Policy With Judgment: Forecast Targeting." *International Journal of Central Banking* 1: 1–54.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society: Series B (Methodological)* 58, no. 1: 267–288.

Wu, J. C., and F. D. Xia. 2020. "Negative Interest Rate Policy and the Yield Curve." *Journal of Applied Econometrics* 35, no. 6: 653–672.

Yu, M., V. Gupta, and M. Kolar. 2019. "Constrained High Dimensional Statistical Inference." arXiv preprint arXiv:1911.07319.

Zhao, P., and B. Yu. 2006. "On Model Selection Consistency of Lasso." *Journal of Machine Learning Research* 7: 2541–2563.

Appendix A

The Test for Forecast Encompassing

We aim to design a test for multiple forecast encompassing that explicitly considers the constraints on the estimated weights and that can deal with high dimensional frameworks.

High dimensionality and constrained parameter space create, however, two problems. First, in the presence of high dimensional nuisance parameters, the maximum likelihood estimator is no longer consistent and, despite it being possible to identify some maximum penalized likelihood estimators that are consistent under some conditions, they may not have a tractable limiting distribution even in the fixed dimensional case (Fu and Knight 2000). This invalidates the standard inferential theory for Wald, Lagrange Multiplier (LM), and Likelihood Ratio (LR) (see for an example Ning and Liu 2017). Second, the nonnegativity and sum to one are inequality constraints that generate a closed and convex *cone* $\boldsymbol{C}$.

We follow the decorrelation procedure of Ning and Liu (2017) and Yu, Gupta, and Kolar (2019) to handle the impact of high dimensional nuisance parameters. This approach offers two main advantages: (a) not relying on post-selection Lee et al. (2016), that is, considering the conditional inference given the event that some covariates have been selected, and (b) the possibility of easily extending the testing procedure under inequality constraints.

A.1 | The Testing Procedure

We now introduce the test for forecasts encompassing. To ease the presentation, we split the testing procedure into two steps. In Algorithm 1, we estimate the ex post optimal combination of forecasts and estimate the decorrelated score function, parameter(s) of interest, and likelihood function. This extends the profile score encountered in low-dimensional problems to high-dimensional frameworks. In Algorithm 2, we form the corresponding test statistics, compute tail probabilities under the null hypothesis and finally perform inference on the combing weight(s).

The key step in Algorithm 1 is to estimate ex post optimal weights and the decorrelation operator $\boldsymbol{W}$. Since both the weights and the nuisance score functions can be high dimensional (have dimension d and d − 1, respectively), we need to impose some assumptions on $\boldsymbol{\beta}$ and $\boldsymbol{W}$ to bound the estimation error when $d > n$. On the one hand, $\boldsymbol{\beta}$ searches for the linear combination of weights that minimizes the variance of the aggregate forecasting error under both the nonnegativity and sum-to-one constraints. When $d > n$, we impose $\boldsymbol{\beta}$ to be sparse. Similarly, when $d > n$, $\boldsymbol{W}$ searches for the best sparse linear combination of the nuisance score functions to approximate the score function.

The key step in Algorithm 2 is to compute the cumulative tail probability of the considered test statistics for a given significance level $\boldsymbol{a}$. To do this, we have to compute the chi-bar-squared weights. When the constrained space is the nonnegative orthant (or more generally when the constraints defining $\boldsymbol{C}$ are linear and independent), Silvapulle and Sen (2004) propose a simple quadratic programming problem to estimate ω (see proposition 3.6.1). When θ is univariate ($r = 1$), the weights are known in analytical form and the chi-bar-squared statistics reduce to

$$\begin{aligned} &\text{Prob}\left[\overline{\chi}^2(\widetilde{\mathscr{F}}_{\theta|\gamma}, \mathbb{R}^{+r}) > c_a\right] = .5\ \text{Prob}\left[\chi^2(0) > c_a\right] \\ &+ .5\ \text{Prob}\left[\chi^2(1) > c_a\right]. \end{aligned} \tag{A1}$$

Algorithm 1 Decorrelation procedure for high dimensional regression frameworks.

1: Estimate ex post optimal weights

$$\hat{\beta}_{\setminus d} = \underset{\beta \in \mathbb{R}^{+(d-1)}}{\arg\min} \left\| V_d - \boldsymbol{V}_{\setminus d} \boldsymbol{\beta}_{\setminus d} \right\|_2^2 , \quad \hat{\beta}_d = 1 - \sum_{i=1}^{d-1} \hat{\beta}_{\setminus d,i} ,$$

2: Define the partition $\beta = (\theta, \gamma)$

3: Estimate the decorrelation operator W_j

$$\widehat{W}_j = \underset{W_j \in \mathbb{R}^{d-1}}{\arg\min} \|V_j - \boldsymbol{V}_{\setminus j} W_j\|_2^2 + \lambda \|W_j\|_1 ,$$

4: When θ is a r-low-dimensional multi-dimensional parameter of interest corresponding to the forecasts in the set I, W can be solved column by column i.e., $\widehat{\boldsymbol{W}} = \left(\widehat{W}_j, \text{ where } j \in I\right)$

5: Estimate the decorrelated score function:

$$\tilde{S}(\theta) = \nabla_\theta \ell(\theta, \hat{\boldsymbol{\gamma}}) - \boldsymbol{W}^T \nabla_\gamma \ell(\theta, \hat{\boldsymbol{\gamma}}) ,$$

6: Estimate the decorrelated parameter of interest:

$$\tilde{\theta} = \hat{\theta} - \widetilde{\mathcal{F}}_{\theta|\gamma}^{-1} \tilde{S}, \quad \text{where} \quad \widetilde{\mathcal{F}}_{\theta|\gamma} = \nabla^2_{\theta\theta} \ell(\hat{\boldsymbol{\beta}}) - \boldsymbol{W}^T \nabla^2_{\theta\gamma} \ell(\hat{\boldsymbol{\beta}}) ,$$

7: Estimate the decorrelated likelihood function:

$$\tilde{\ell}(\boldsymbol{\theta}) = \ell(\theta, \hat{\boldsymbol{\gamma}} - \boldsymbol{W}^T(\theta - \hat{\theta})) .$$

Algorithm 2 One-sided test for forecasts encompassing.

1: Form the decorrelated score test statistics:
$T_S = \left[\tilde{S}(\hat{\theta}_{H_0}) - \tilde{S}(\hat{\theta}_{H_1})\right]^T \widetilde{\mathcal{F}}_{\theta|\gamma}^{-1} \left[\tilde{S}(\hat{\theta}_{H_0}) - \tilde{S}(\hat{\theta}_{H_1})\right]$.

2: Form the decorrelated likelihood ratio (LR) test statistics:
$T_{LR} = 2n\left(\tilde{\ell}(\hat{\boldsymbol{\beta}}_{H_0}) - \tilde{\ell}(\hat{\boldsymbol{\beta}}_{H_1})\right)$,

3: Form the decorrelated Wald test statistics:
$T_W = \left[\tilde{\theta} - \hat{\theta}_{H_0}\right]^T \widetilde{\mathcal{F}}_{\theta|\gamma} \left[\tilde{\theta} - \hat{\theta}_{H_0}\right] - \left[\tilde{\theta} - \hat{\theta}_{H_1}\right]^T \widetilde{\mathcal{F}}_{\theta|\gamma} \left[\tilde{\theta} - \hat{\theta}_{H_1}\right]$.

4: Compute the set of mixing weights $\omega_i(r, \widetilde{\mathcal{F}}_{\theta|\gamma}, \mathbb{R}^{+r})$ for the $\overline{\chi}^2$ statistics.

5: Given a significance level a and its corresponding critical value c_a, compute cumulative tail probability :

$$\text{Prob}\left[\overline{\chi}^2(\widetilde{\mathcal{F}}_{\theta|\gamma}, \mathbb{R}^{+r}) > c_a\right] = \sum_{i=0}^{r} \omega_i \text{Prob}\left[\chi^2(i) > c_a\right] .$$

6: Reject H_0 if

$$\text{Prob}\left[T_S > c\right] > a , \quad \text{Prob}\left[T_W > c\right] > a, \quad \text{Prob}\left[T_{LR} > c\right] > a .$$

To ease presentation and simplify notation, we focus on the quasi-likelihood framework and on the univariate parameter of interest θ ($r = 1$) corresponding to the j^{th} forecast. Nevertheless, the decorrelated score function can be defined for a multi-dimensional parameter of interest and also when $\boldsymbol{\beta}$ minimizes a loss function other than the negative log-likelihood (see section 3 and supplementary material of Ning and Liu 2017, respectively).

A.1.1 | Assumptions of the Test When $d > n$

In this section, we describe the high-level assumptions required to obtain weak convergence of test under the null hypothesis in the high dimensional framework. We closely follow Ning and Liu (2017) and Yu, Gupta, and Kolar (2019) and refer to their papers for proof of the results in the case of linear regression models. These conditions can be classified into three main categories: (a) consistency conditions for initial parameter estimation (Assumptions 1, 2, and 3); (b) concentration of the gradient and Hessian matrix (Assumption 4); (c) local smoothness on the loss function (Assumption 5). All these assumptions are fairly standard in the high dimensional statistics literature and are known to hold for the Lasso and nonnegative least squares with sum to one constraint on the weights (Fan et al. 2012).

Assumption 1. Score condition. The expected value of the score function at the true $\boldsymbol{\beta}^*$ is equal to zero, that is, $\nabla \ell(\boldsymbol{\beta}^*) = 0$.

Assumption 2. Restricted eigenvalue condition for the covariance matrix. Given the positive definite matrix $\mathcal{F}$ and $\nabla \ell^2(\boldsymbol{\beta})$, let $\mathcal{S} = supp(\boldsymbol{\beta}^*) \cup supp(\boldsymbol{W}^*)$. There exists a universal constant $c, k > 0$ such that the quadratic form $\boldsymbol{v}^t \mathcal{F} \boldsymbol{v} \geq k \|\boldsymbol{v}\|_2^2$, and $\boldsymbol{v}^T \nabla \ell^2(\boldsymbol{\beta}) \boldsymbol{v}$, for any $\boldsymbol{v}$ in the cone C, that is, $\forall \boldsymbol{v} \in C(\mathcal{S}) = \left\{\boldsymbol{v}\colon \|\boldsymbol{v}_{\mathcal{S}^c} \leq c \boldsymbol{v}_{\mathcal{S}}\|\right\}$.

Assumption 3. Consistency conditions for initial parameter estimation. As $n \to \infty$, for some sequences $\nu_1(n)$ and $\nu_2(n)$ converging to 0, the following holds:

$$\lim_{n\to\infty} \text{Prob}_{\boldsymbol{\beta}^*}\left[\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 \lesssim \nu_1(n)\right] = 1 , \text{ and}$$

$$\lim_{n\to\infty} \text{Prob}_{\boldsymbol{\beta}^*}\left[\|\widehat{\boldsymbol{W}} - \boldsymbol{W}^*\|_1 \lesssim \nu_2(n)\right] = 1 .$$

In lemma 3.1 of the paper, D3 and D4 of the supplementary material, Ning and Liu (2017) show that this assumption holds when considering the Lasso estimator of $\boldsymbol{\beta}$ and $\boldsymbol{W}$ in a linear regression model. In this case with high probability, we have the following

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\|_1 = \mathcal{O}\left(s^* \sqrt{\frac{\log(d)}{n}}\right) ,$$

$$\|\widehat{\boldsymbol{W}}_j - \boldsymbol{W}_j^*\|_1 = \mathcal{O}\left(s' \sqrt{\frac{\log(d)}{n}}\right) \text{with } j \in I ,$$

where $\boldsymbol{\beta}$ and $\boldsymbol{W}$ are defined in Algorithm 1 and $s^* = \|\boldsymbol{\beta}^*\|_0$ and $s' = \|\boldsymbol{W}^*\|_0$, that is, the number of non zero elements in $\boldsymbol{\beta}^*$ and $\boldsymbol{W}^*$, respectively.[28]

Assumption 4. Concentration of the gradient and Hessian. Let $v^* = (1, -\boldsymbol{W}^{T*})$ and assume that the largest element of the score function is bounded, that is, $\| \nabla \ell(\boldsymbol{\beta}^*)\|_\infty = \mathcal{O}\left(\sqrt{\frac{log(d)}{n}}\right)$ and

$$\| v^{T*}\nabla^2\ell(\boldsymbol{\beta}^* - \mathbb{E}_{\boldsymbol{\beta}^*}\left[v^{T*}\nabla^2\ell(\boldsymbol{\beta}^*\right]\|_\infty = \mathcal{O}\left(\sqrt{\frac{\log(d)}{n}}\right).$$

Assumption 5. Local smoothness of the loss function For some constant L, the Hessian matrix $\nabla^2\ell(\boldsymbol{\beta})$ is Lipschitz continuous:

$$\| \nabla^2\ell(\boldsymbol{\beta}_1) - \nabla^2\ell(\boldsymbol{\beta}_2)\|_\infty \leq L \| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|_1 .$$

Lemma 1. Central limit theorem (CLT) for the linear combination of the score function. Let $\boldsymbol{\Sigma}^* = \lim_{n\to\infty} Var_{\boldsymbol{\beta}^*}\left[\sqrt{n}\nabla \updownarrow(\boldsymbol{\beta}^*)\right]$. Under Assumptions 1–5, Ning and Liu (2017) and Yu, Gupta, and Kolar (2019) proved that[29]

$$\sqrt{n}\boldsymbol{v}^{T*}\nabla\ell(\boldsymbol{\beta}^*)\sqrt{\sigma^*}^{-1} \rightsquigarrow \mathcal{N}(0,\ 1),\ \text{where}\ \sigma^* = \boldsymbol{v}^{T*}\boldsymbol{\Sigma}^*\boldsymbol{v}^* .$$

Appendix B

Size and Power Analysis

In this section, we study the performance of the proposed test both in terms of empirical size and power. Similarly to Busetti and Marcucci (2013), we design a Monte Carlo experiment where we model the j^{th} forecast using a single factor model

$$f_i = x_i + \epsilon_{i,j},\ \text{for}\ i = 1,2,\ldots n \tag{B1}$$

and allow for $d > n$. The factor f_i is an i.i.d. $\mathcal{N}(0,1)$ and $\epsilon_{i,j}$ is an idiosyncratic and potentially auto-correlated error term, i.e., $\epsilon_{i,j} = \rho\epsilon_{i-1,j} + z_j, z_j \sim \mathcal{N}(0, \sigma_z)$. The i^{th} realization of the variable of interest is obtained from

$$y_i = \boldsymbol{\beta}^T\boldsymbol{f}_i + u_i . \tag{B2}$$

The systematic error term u_i permits us to study the behavior of the test even in the presence of auto-correlated residuals, i.e., $u_i = \rho u_{i-1} + z_i, z_i \sim \mathcal{N}(0, \sigma_z)$. $\boldsymbol{\beta}$ is the vector weights that lies in the *d-1* dimensional unit simplex. We partition the vector of weights $\boldsymbol{\beta}$ into two sub-vectors, that is, $\boldsymbol{\beta} = (\theta, \boldsymbol{\gamma})$, θ is the parameter of interest while $\boldsymbol{\gamma}$ is a sparse vector of nuisance parameters with d non-zero components. Indeed, we require $\boldsymbol{\gamma}$ to be sparse to ensure the consistency conditions on $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{W}}$ when the problem is high-dimensional.

We consider several schemes to assign the weights to the non-zero element i of $\boldsymbol{\gamma}$: (a) equal weight, that is, $\boldsymbol{\gamma}_i = \frac{1-\theta}{d}$; (b) random, that is, $\boldsymbol{\gamma}_i = \frac{r}{k}$, where $r \sim \mathcal{U}(0,1)$ and k is a normalization constant such that $\theta + \boldsymbol{\gamma} = 1$. To avoid repetition, we report the results for the scheme (a) with $m = 9$, $d = 5$, $n = \{28, 56, 112\}$, $\rho = \{0, 0.25, 0.5, 0.75\}$ and $\sigma_z = 0.5$ and $\sigma_u = 0.25$.[30]

Given n realizations of the target and m forecasts, Tables A1, A2, and A3 report the results of the simulation study for a nominal size of 5% and 10,000 repetitions. We confirm the usual results about the size-adjusted power curve. Power increases when $n\uparrow$ and when $\theta \to 1$. We also expect that when $\rho\uparrow$, the size-adjusted power $\downarrow$. Indeed, we have specified the dynamics of the ϵ_j as an autoregressive process of order 1—AR(1)—and, keeping σ_z constant, the long-term variance of ϵ_j to tend to infinity as $\rho \to 1$. Empirical and size-adjusted rejection probabilities for the analysis of the empirical size and size-adjusted power in a large sample and in a high dimensional framework are reported in Tables A4 and A5.

TABLE A1 | Empirical rejection rate (Empirical Rejection) and size-adjusted (SA - Power) rejection probabilities for analyzing the empirical size and size-adjusted power when considering the nonnegativity constraint (C) or not (UC) the testing procedure.

	Distance from the null						
$n = 28, \rho = 0$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.059	0.062	0.092	0.194	0.523	0.687
Empirical Rejection - C	0.081	0.086	0.097	0.131	0.244	0.582	0.743
SA - Power - UC	0.050	0.054	0.056	0.078	0.158	0.451	0.615
Empirical Rejection - UC	0.122	0.121	0.132	0.162	0.277	0.608	0.755
$n = 56, \rho = 0$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.049	0.072	0.113	0.317	0.827	0.949
Empirical Rejection - C	0.068	0.067	0.088	0.140	0.358	0.856	0.961
SA - Power - UC	0.050	0.050	0.067	0.109	0.297	0.812	0.943
Empirical Rejection - UC	0.079	0.077	0.098	0.151	0.372	0.859	0.964
$n = 112, \rho = 0$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.058	0.089	0.198	0.593	0.989	1.000
Empirical Rejection - C	0.055	0.062	0.095	0.212	0.610	0.990	1.000
SA - Power - UC	0.050	0.056	0.088	0.193	0.586	0.988	1.000
Empirical Rejection - UC	0.061	0.066	0.099	0.218	0.616	0.990	1.000

Note: We estimate the empirical size by generating the optimal forecast under the null hypothesis, that is, by enforcing the weight to be tested to be zero in the population. We conduct the power analysis by generating the optimal forecast under the alternative for a departure from the null of $\beta = 0$, that is, by imposing the population weight of the forecast to be tested equal to 0.01, 0.025, 0.05, 0.10, 0.20, and 0.25.

TABLE A2 | Empirical and size-adjusted rejection probabilities for the analysis of the empirical size and size-adjusted power.

	Distance from the null						
$n = 28, \rho = 0.25$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.052	0.068	0.091	0.176	0.500	0.660
Empirical Rejection - C	0.081	0.085	0.102	0.125	0.232	0.574	0.726
SA - Power - UC	0.050	0.052	0.066	0.080	0.146	0.444	0.603
Empirical Rejection - UC	0.117	0.119	0.138	0.157	0.262	0.597	0.745
SA - Power - C - NW	0.050	0.054	0.074	0.098	0.167	0.459	0.608
Empirical Rejection - C - NW	0.075	0.079	0.101	0.127	0.216	0.529	0.679
SA - Power - UC - NW	0.050	0.051	0.067	0.065	0.116	0.331	0.472
Empirical Rejection - UC - NW	0.140	0.149	0.161	0.174	0.259	0.565	0.706
$n = 56, \rho = 0.25$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.054	0.064	0.121	0.306	0.810	0.939
Empirical Rejection - C	0.066	0.072	0.084	0.142	0.349	0.836	0.955
SA - Power - UC	0.050	0.054	0.063	0.113	0.293	0.793	0.932
Empirical Rejection - UC	0.080	0.083	0.095	0.154	0.362	0.843	0.957
SA - Power - C - NW	0.050	0.058	0.073	0.126	0.322	0.780	0.925
Empirical Rejection - C - NW	0.064	0.072	0.090	0.147	0.356	0.807	0.937
SA - Power - UC - NW	0.050	0.051	0.056	0.097	0.249	0.697	0.876
Empirical Rejection - UC - NW	0.100	0.096	0.110	0.163	0.369	0.818	0.940
$n = 112, \rho = 0.25$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.060	0.079	0.185	0.565	0.984	0.999
Empirical Rejection - C	0.057	0.064	0.086	0.195	0.581	0.986	0.999
SA - Power - UC	0.050	0.058	0.073	0.179	0.549	0.983	0.998
Empirical Rejection - UC	0.062	0.068	0.091	0.200	0.586	0.986	0.999
SA - Power - C - NW	0.050	0.061	0.089	0.190	0.545	0.979	0.998
Empirical Rejection - C - NW	0.060	0.069	0.102	0.206	0.571	0.983	0.998
SA - Power - UC - NW	0.050	0.060	0.079	0.166	0.507	0.973	0.997
Empirical Rejection - UC - NW	0.075	0.081	0.108	0.215	0.579	0.983	0.998

Note: We estimate the empirical size by generating the optimal forecast under the null hypothesis, that is, by enforcing the weight to be tested to be zero in the population. We conduct the power analysis by generating the optimal forecast under the alternative for a departure from the null of $\beta = 0$, that is, by imposing the population weight of the forecast to be tested equal to 0.01, 0.025, 0.05, 0.10, 0.20, and 0.25. To handle the autocorrelation of the forecast errors, we also compute the score statistics using Newey and West (1987) HAC covariance matrix.

TABLE A3 | Empirical and size-adjusted rejection probabilities for the analysis of the empirical size and size-adjusted power in a large sample and in a high dimensional framework.

	Distance from the null						
$n = 28, \rho = 0.5$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.055	0.058	0.081	0.156	0.430	0.573
Empirical Rejection - C	0.081	0.081	0.090	0.115	0.203	0.509	0.643
SA - Power - UC	0.050	0.054	0.053	0.073	0.137	0.375	0.518
Empirical Rejection - UC	0.115	0.116	0.117	0.143	0.235	0.532	0.666
SA - Power - C - NW	0.050	0.055	0.063	0.082	0.156	0.392	0.529
Empirical Rejection - C - NW	0.074	0.076	0.089	0.109	0.199	0.465	0.602
SA - Power - UC - NW	0.050	0.053	0.050	0.066	0.111	0.289	0.407
Empirical Rejection - UC - NW	0.134	0.131	0.136	0.163	0.241	0.509	0.639
$n = 56, \rho = 0.5$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.048	0.058	0.094	0.254	0.719	0.869
Empirical Rejection - C	0.070	0.069	0.079	0.120	0.297	0.757	0.896
SA - Power - UC	0.050	0.046	0.054	0.088	0.239	0.693	0.854
Empirical Rejection - UC	0.082	0.077	0.089	0.134	0.308	0.766	0.899
SA - Power - C - NW	0.050	0.059	0.066	0.101	0.260	0.697	0.847
Empirical Rejection - C - NW	0.068	0.073	0.085	0.123	0.300	0.739	0.872
SA - Power - UC - NW	0.050	0.052	0.050	0.074	0.194	0.607	0.780
Empirical Rejection - UC - NW	0.100	0.098	0.105	0.148	0.316	0.750	0.879
$n = 112, \rho = 0.5$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.051	0.076	0.152	0.462	0.953	0.994
Empirical Rejection - C	0.059	0.061	0.088	0.170	0.491	0.960	0.995
SA - Power - UC	0.050	0.049	0.073	0.148	0.449	0.950	0.994
Empirical Rejection - UC	0.064	0.065	0.093	0.177	0.496	0.961	0.995
SA - Power - C - NW	0.050	0.054	0.084	0.168	0.468	0.948	0.993
Empirical Rejection - C - NW	0.059	0.063	0.093	0.185	0.494	0.955	0.993
SA - Power - UC - NW	0.050	0.052	0.074	0.147	0.430	0.935	0.989
Empirical Rejection - UC - NW	0.076	0.077	0.103	0.193	0.501	0.956	0.993

Note: We estimate the empirical size by generating the optimal forecast under the null hypothesis, that is, by enforcing the weight to be tested to be zero in the population. We conduct the power analysis by generating the optimal forecast under the alternative for a departure from the null of $\beta = 0$, that is, by imposing the population weight of the forecast to be tested equal to 0.01, 0.025, 0.05, 0.10, 0.20, and 0.25. To handle the autocorrelation of the forecast errors, we also compute the score statistics using Newey and West (1987) HAC covariance matrix.

TABLE A4 | Empirical and size-adjusted rejection probabilities for the analysis of the empirical size and size-adjusted power in a large sample and in a high dimensional framework.

	Distance from the null						
$n = 28$, $\rho = 0.75$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.052	0.059	0.070	0.122	0.299	0.431
Empirical Rejection - C	0.075	0.077	0.091	0.105	0.163	0.363	0.494
SA - Power - UC	0.050	0.055	0.061	0.067	0.103	0.268	0.383
Empirical Rejection - UC	0.111	0.110	0.120	0.132	0.191	0.391	0.516
SA - Power - C - NW	0.050	0.059	0.071	0.087	0.142	0.320	0.423
Empirical Rejection - C - NW	0.060	0.064	0.078	0.094	0.151	0.337	0.438
SA - Power - UC - NW	0.050	0.054	0.051	0.056	0.084	0.213	0.303
Empirical Rejection - UC - NW	0.117	0.123	0.126	0.137	0.189	0.376	0.485
$n = 56$, $\rho = 0.75$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.052	0.059	0.089	0.181	0.526	0.707
Empirical Rejection - C	0.061	0.064	0.070	0.104	0.203	0.562	0.732
SA - Power - UC	0.050	0.051	0.057	0.088	0.173	0.509	0.691
Empirical Rejection - UC	0.071	0.075	0.081	0.118	0.216	0.576	0.740
SA - Power - C - NW	0.050	0.058	0.073	0.104	0.199	0.541	0.701
Empirical Rejection - C - NW	0.056	0.063	0.081	0.112	0.212	0.557	0.712
SA - Power - UC - NW	0.050	0.052	0.063	0.083	0.162	0.469	0.631
Empirical Rejection - UC - NW	0.084	0.090	0.100	0.131	0.232	0.570	0.723
$n = 112$, $\rho = 0.75$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.051	0.068	0.109	0.312	0.814	0.926
Empirical Rejection - C	0.060	0.061	0.078	0.124	0.342	0.836	0.937
SA - Power - UC	0.050	0.050	0.067	0.107	0.306	0.805	0.922
Empirical Rejection - UC	0.065	0.065	0.083	0.128	0.346	0.839	0.938
SA - Power - C - NW	0.050	0.056	0.074	0.125	0.335	0.805	0.920
Empirical Rejection - C - NW	0.058	0.064	0.084	0.138	0.359	0.823	0.926
SA - Power - UC - NW	0.050	0.051	0.069	0.111	0.302	0.777	0.904
Empirical Rejection - UC - NW	0.075	0.074	0.093	0.148	0.368	0.827	0.928

Note: We estimate the empirical size by generating the optimal forecast under the null hypothesis, that is, by enforcing the weight to be tested to be zero in the population. We conduct the power analysis by generating the optimal forecast under the alternative for a departure from the null of $\beta = 0$, that is, by imposing the population weight of the forecast to be tested equal to 0.01, 0.025, 0.05, 0.10, 0.20, and 0.25. To handle the autocorrelation of the forecast errors, we also compute the score statistics using Newey and West (1987) HAC covariance matrix.

TABLE A5 | Empirical and size-adjusted rejection probabilities for the analysis of the empirical size and size-adjusted power in a large sample and in a high dimensional framework.

	Distance from the null						
	0 - size	**0.01**	**0.025**	**0.05**	**0.10**	**0.20**	**0.25**
$n = 1000$, d $= 20$, $\rho = 0$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.100	0.370	0.898	1.000	1.000	1.000
Empirical Rejection - C	0.051	0.103	0.374	0.900	1.000	1.000	1.000
SA - Power - UC	0.050	0.097	0.362	0.894	1.000	1.000	1.000
Empirical Rejection - UC	0.055	0.104	0.380	0.902	1.000	1.000	1.000
$n = 40$, d $= 80$, $\rho = 0$	0	0.01	0.025	0.05	0.1	0.2	0.25
SA - Power - C	0.050	0.047	0.051	0.071	0.127	0.341	0.461
Empirical Rejection - C	0.054	0.051	0.058	0.076	0.138	0.360	0.479
SA - Power - UC	0.050	0.046	0.051	0.057	0.096	0.242	0.349
Empirical Rejection - UC	0.035	0.033	0.037	0.041	0.071	0.198	0.290

Note: We study the convergence of the empirical to the theoretical 5% size for a sample size of 1000 observations and the behavior of the proposed test in the high dimensional framework with 40 observations and 80 candidate forecasts. In this latter case, we estimate the unconstrained weights using Lasso. The increase of the sample size was necessary to estimate the penalty intensity via 5-fold cross-validation. We estimate the empirical size by generating the optimal forecast under the null hypothesis, that is, by enforcing the weight to be tested to be zero in population. We conduct the power analysis by generating the optimal forecast under the alternative for a departure from the null of $\beta = 0$, that is, by imposing the population weight of the forecast to be tested equal to 0.01, 0.025, 0.05, 0.10, 0.20, and 0.25.

Appendix C

On the Empirical Distribution of ϵ

We analyze the distribution of the residual forecast error, denoted by ϵ, which is implied by the ex-post optimal combinations of forecasts. In Section 3.2, we use the Gaussian quasi-likelihood to derive the distribution of the test under the null hypothesis. Here, we present the descriptive statistics of ϵ in Table A6 and compare our empirical distribution to the Normal density using the Kolmogorov-Smirnov test. The p-value of 0.319 suggests that we cannot reject the null hypothesis of no difference between the two distributions at the usual 90%, 95%, and 99% confidence levels.

We present the empirical distribution of ϵ based on our sample of 56 realizations spanning from 2009 to 2022 in Figure A1. This distribution is compared with the Normal density. The central tendency of the distribution is around 0, showing no apparent skewness. However, there is some indication of leptokurtic tails. To delve deeper into this observation, we introduce a quantile-on-quantile (QQ) plot in Figures A2 and A3. This plot compares the empirical quantiles of ϵ with the theoretical quantiles of the Normal density. In the context of a sample comprising N observations, we construct pointwise confidence bands for the sample p-quantile q^p. These bands are derived from the Normal distribution centered around q^p with a variance of $\frac{p(1-p)}{N\left[\phi(\Phi^{-1}(p))\right]^2}$, where ϕ and Φ denote the probability density function (pdf) and cumulative distribution function (cdf) of a standard Normal variable, respectively (Hyndman and Fan 1996).

Figure A2 reveals some indications of data tails. To address this, we examine a QQ plot for the sample excluding the years 2020, 2021, and 2022 in Figure A3. The presence of fat tails diminishes substantially, suggesting that leptokurtosis was primarily driven by observations in 2020 and 2021, particularly in Q1 (observations 45 and 49). These observations coincide with the inflation forecasts made at the onset of the COVID-19 pandemic and the Russian invasion of Ukraine. In this context, the theoretical and empirical quantiles align more closely, diminishing concerns regarding leptokurtic tails. Additionally, we emphasize that any residual evidence of leptokurtosis may be attributed to the small sample size.

TABLE A6 | Descriptive statistics of the prediction errors of the average optimal combinations of forecasts.

	Mean	Sd	Min	Max	Range	Skew	Kurtosis	
	−0.12	0.40	−1.39	1.18	2.57	−0.1	2.5	
Percentiles								
2.50%	5%	10%	25%	50%	75%	90%	95%	97.50%
−0.950	−0.713	−0.558	−0.256	−0.064	0.050	0.182	0.496	0.658
Kolmogorov-Smirnov test p-value:							0.319	

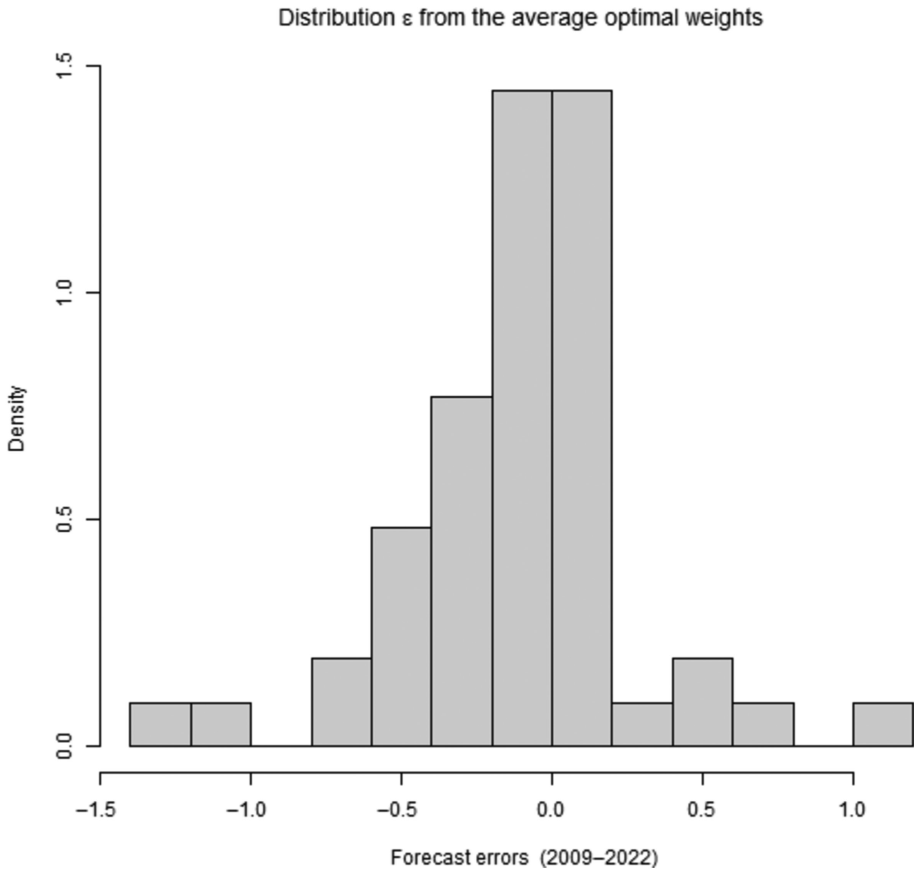

FIGURE A1 | Distribution the realized ϵ associated with the average (Spec.1, 2, and 3) optimal combining weights.

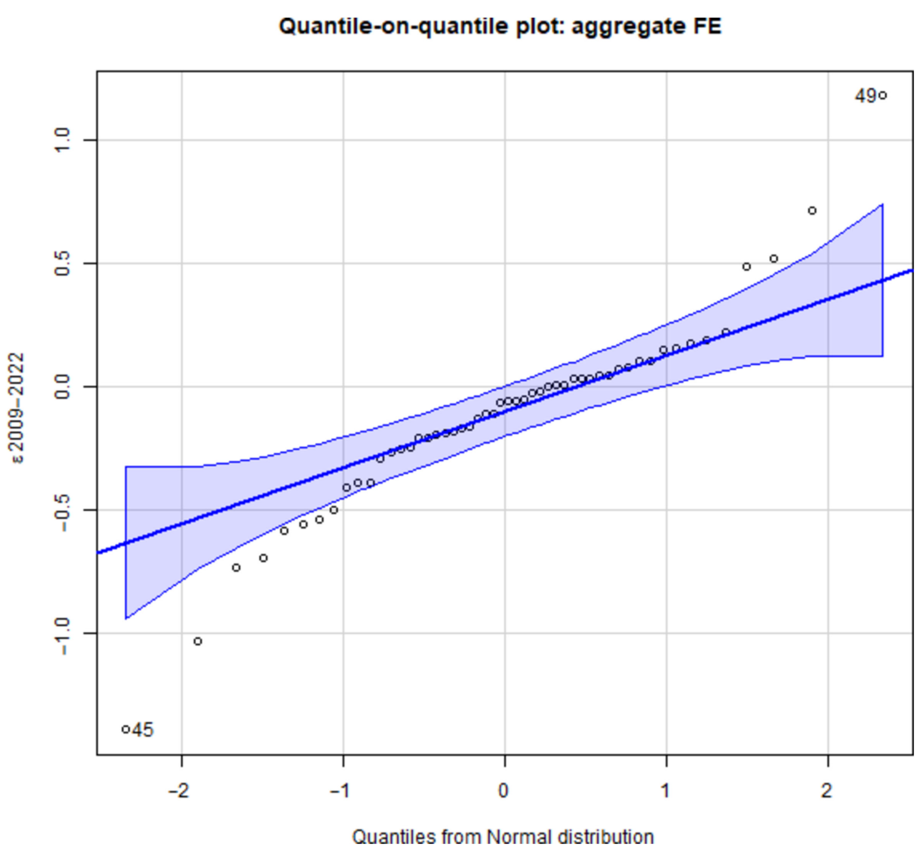

FIGURE A2 | Quantile-on-quantile plot of the realized ϵ associated with the average optimal combining weights with the theoretical quantiles from the normal distribution with 99% confidence bands.

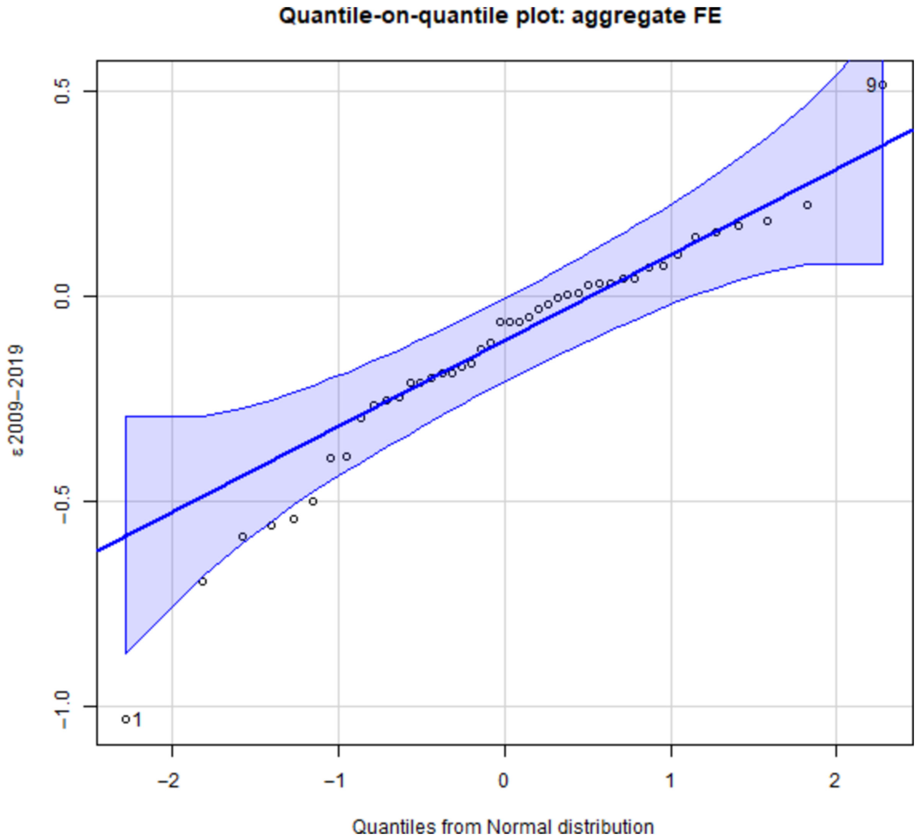

FIGURE A3 | Quantile-on-quantile plot of the realized ϵ as in Figure A2 but when excluding the Covid period (2020–2021).

Appendix D

Monthly Estimates of the Combining Weights

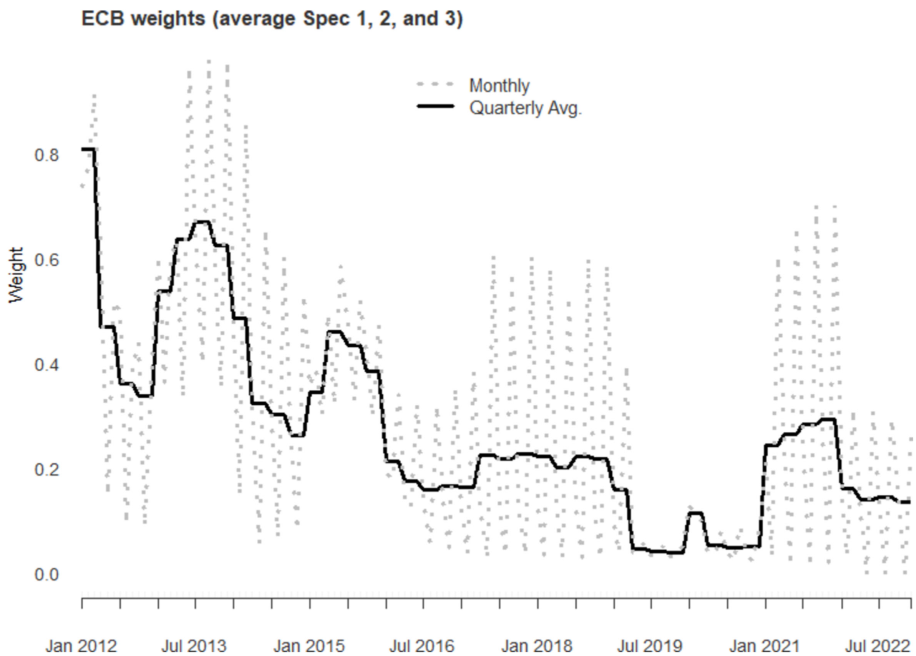

FIGURE A4 | Monthly and quarterly (after averaging) estimates of the ECB optimal weights. The series display the weights obtained by averaging the POET, Sample, and LW estimates.

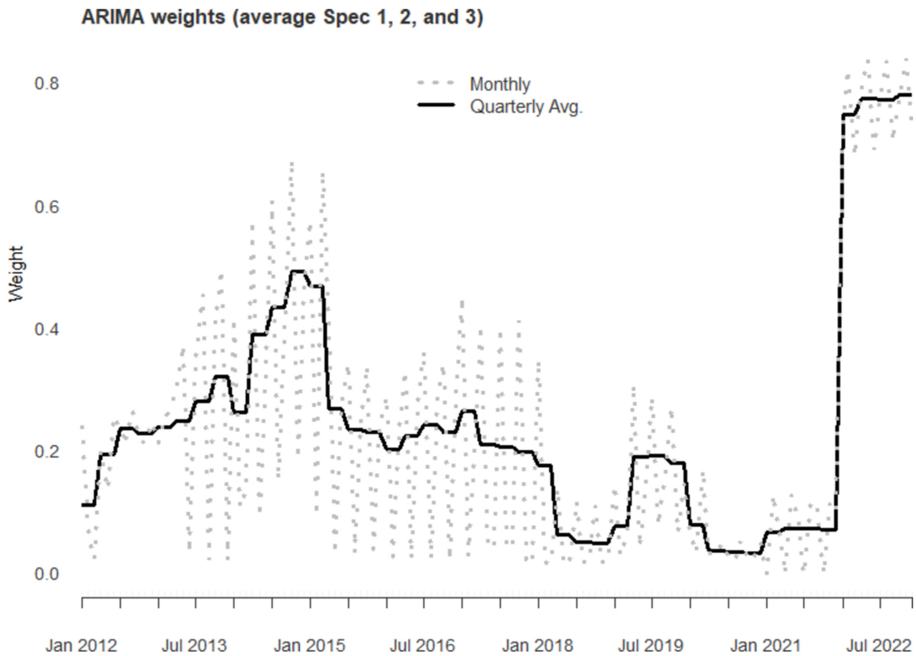

FIGURE A5 | Monthly and quarterly (after averaging) estimates of the ARIMA optimal weights. The series display the weights obtained by averaging the POET, Sample, and LW estimates.

Appendix E

ECB Weight on Subset of the Pool of Forecasters

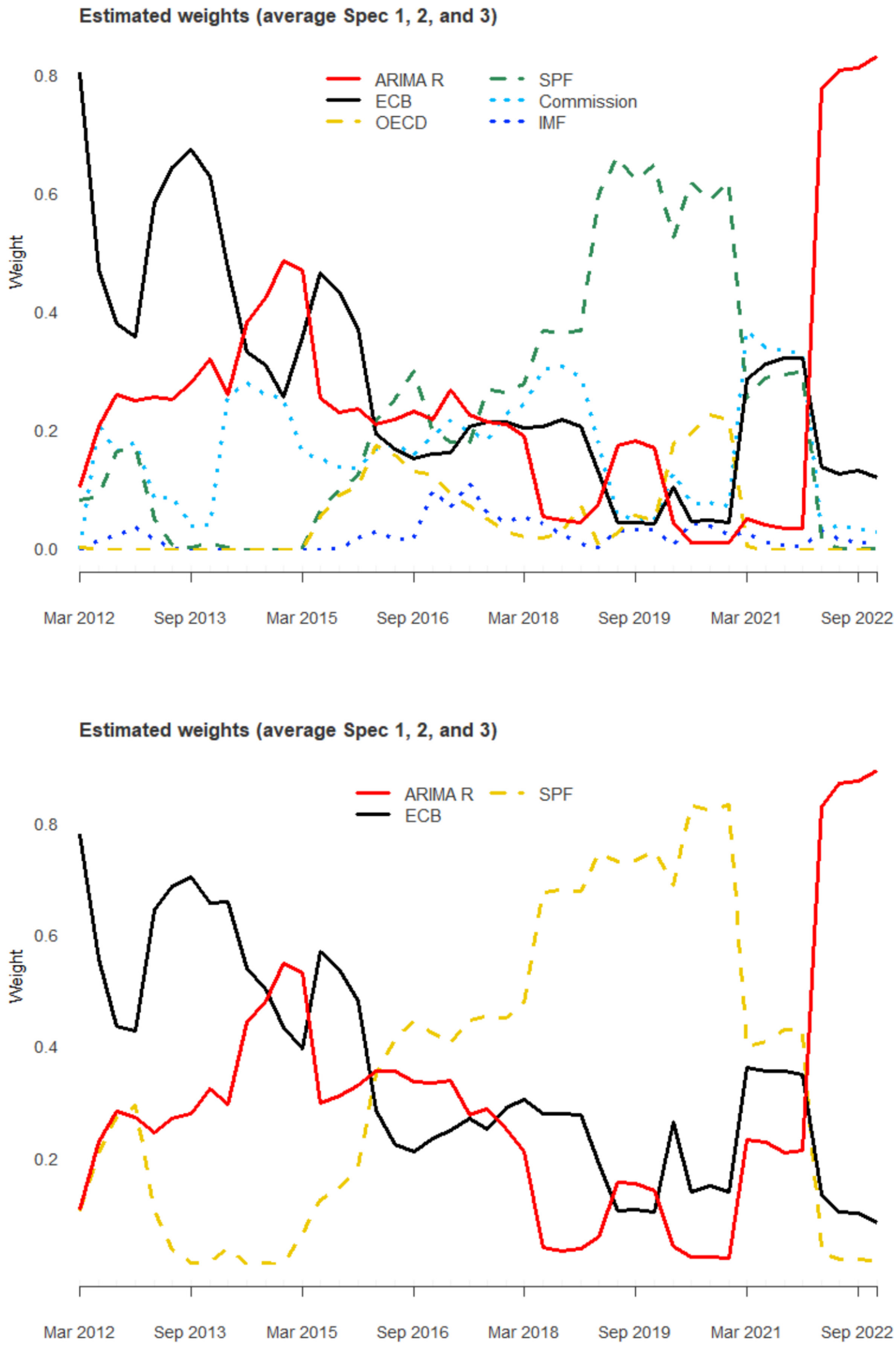

FIGURE A6 | Dynamics of the average ECB weight across specifications for different subsets of forecasters.

Appendix F

Sensitivity of the ECB Weight to 16 and 20 Quarters Rolling Window

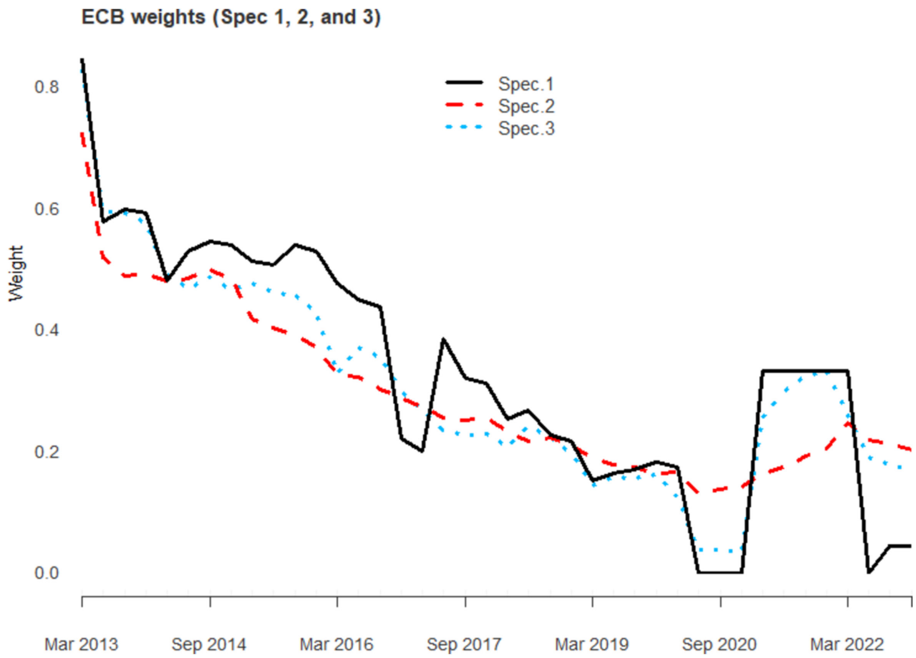

FIGURE A7 | ECB optimal combining weights estimated with 16 quarters rolling window.

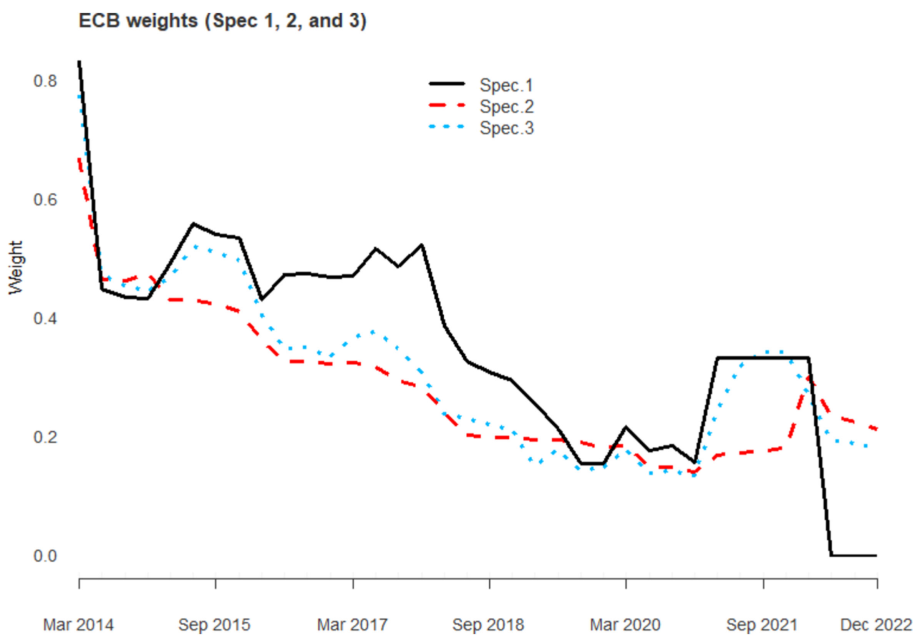

FIGURE A8 | ECB optimal combining weights estimated with 20 quarters rolling window.

Appendix G

Beta Regression Framework

The density function of a beta distribution is given by

$$f(y,p,q)=\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1},\ \text{with } 0<y<1\ , \tag{G1}$$

with $p,q>0$ and $\Gamma(\cdot)$ being the gamma function. However, the beta distribution can be conveniently re-parameterized in terms of $0<\mu<1$ and $\phi>0$ that are related to the mean and variance of the dependent variable y,

$$E[y_i]=\mu_i\ \text{ and }\ Var[y_i]=\frac{\mu_i(1-\mu_i)}{1+\phi}\ . \tag{G2}$$

In this framework, the log-likelihood function becomes

$$\ell(\mu,\phi)=\sum_{i=1}^{n}\ell_i(\mu_i,\phi)\ , \tag{G3}$$

where

$$\ell_i(\mu_i,\phi)=log[\Gamma(\phi)]-log\left[\Gamma\left(\mu_i\phi\right)\right]-log\left\{\Gamma\left[(1-\mu_i)\phi\right]\right\} + \left(\mu_i\phi-1\right)log\left(y_i\right)+\left[\left(1-\mu_i\right)\phi-1\right]log(1)y_i)\ .$$

Using this notation, the standard beta regression model assumes (a) independent realizations of $y_i\sim\mathcal{B}(\mu_i,\phi_i)$, (b) a linear regression model for the transform of the mean parameter, and (c) a constant precision parameter ϕ. Fixing the precision parameter makes the beta regression framework more parsimonious at the cost of lower flexibility than the variable dispersion beta regression models featured in Gambetti, Gauthier, and Vrins (2019). However, this is a price we have to pay considering that we count only 38 observations in our sample. Nevertheless, we can easily notice that this model can still naturally accommodate heteroscedasticity because $Var[y]$ is also a function of μ_i. Formally,

$$g_1\left(\mu_i\right)=x_i^T\beta\ \text{ and }\ g_2\left(\phi_i\right)=\phi\ , \tag{G4}$$

with $g_1(\cdot):(0,1)\rightarrow\mathbb{R}$ and $g_2(\cdot):(0,\infty)\rightarrow\mathbb{R}$, with x_i being the vector of regressors, β conformable vectors of regression coefficients, and $g_1(\cdot)$ and $g_2(\cdot)$ being strictly increasing and twice-differentiable link functions. We consider a *logit* link function for g_1 and a constant g_2. This results in

$$\mu_i=\frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\ \text{and}\ \phi_i=\phi\ . \tag{G5}$$

Appendix H

Bootstrap Procedure

In our framework, we consider the bootstrapping procedure to take into account also unknown forms of heteroskedasticity that could emerge in the dynamics of the estimated weights, we proposed the wild bootstrapped procedure as in Mammen (1993). Specifically,

1. Estimate the standard linear regression or beta regression parameters via maximum likelihood as described in Section 4.2 (after retrieving the ex post optimal combination weights as outlined in Section 3.2), and the corresponding residuals u_t^i for each $i=1,\ \ldots\ ,n$.
2. Draw with replacement the sequence of errors $\{\tilde{u}_i=k_i\hat{u}_i\}_{i=1}^n$ with

$$k_i=\begin{cases}\frac{1+\sqrt{5}}{2}, & \text{with probability } p=\frac{\sqrt{5}-1}{2\sqrt{5}}\\ \frac{1-\sqrt{5}}{2}, & \text{with probability } 1-p.\end{cases}$$

   as in Mammen (1993) with $\{\hat{u}_i\}_{i=1}^n$ being the estimated errors in the beta regression.
3. Generate the bootstrapped dependent variable using $\{\tilde{y}_i=\hat{y}_i+k_i\}_{i=1}^n$, with $\{\hat{y}_i\}_{i=1}^n$ being the fitted values of the dependent variable in the beta regression. We map negative values or weights greater than one of the bootstrapped samples $\{\tilde{y}_i\}_{i=1}^n$ to zero and one, respectively. Following Smithson and Verkuilen (2006), we consider the transformation $\frac{y(n-1)+0.5}{n}$ where y is the dependent variable and n is the sample size to adapt the beta regression framework to dependent variables assuming also the extremes 0 and 1.
4. Estimate the beta regression on the bootstrapped data sample.
5. Repeat 2–4 a large number of times (in this work, we repeat the procedure 10,000 times) and then extract the quantiles needed.

Appendix I

Additional Results on the Determinants of the Weights

TABLE A7 | Determinants of the weights constructed using the Spec. 1 weights.

(a)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.874	0.084	***	***	2.069	0.587	***	***
Dif. 2%	−0.055	0.014	***	***	−0.231	0.083	***	***
EPU	−0.001	0.000	***	***	−0.008	0.002	***	***
S	0.064	0.007	***	***	0.357	0.063	***	***
Inf. Sur.	−0.035	0.010	***	***	−0.152	0.087	*	***
Glob.	−0.003	0.007			0.028	0.047		
Frag.	−0.031	0.016	*	*	−0.171	0.085	**	*
R^2	0.833				0.6237			

(b)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.805	0.290	***	***	4.307	3.440		
Dif. 2%	−0.062	0.025	**	***	−0.298	0.179	*	*
Dis.	0.018	0.044			−0.217	0.407		
EPU	−0.001	0.000	***	***	−0.007	0.003	**	***
S	0.065	0.010	***	***	0.406	0.100	***	***
APP	−0.010	0.020			−0.166	0.155		
Inf. Sur.	−0.039	0.015	**	***	−0.211	0.126	*	*
Glob.	0.000	0.009			0.016	0.047		
Frag.	−0.036	0.014	**	*	−0.175	0.083	**	**
Oil	0.005	0.055			−1.589	1.957		
Spread	0.089	0.066			0.339	0.436		
d_2	−0.086	0.047	*		−0.445	0.321		
d_3	−0.043	0.040			−0.224	0.304		
d_4	−0.046	0.054			−0.309	0.395		
R^2	0.857				0.5523			

Note: The first column displays the determinants; b_0 is the intercept. LM and Beta report the estimated coefficients using the linear regression model and the beta regression. We report the adjusted R^2 for the LM and adjusted pseudo R^2 for the Beta. We display Newey-West standard errors in column SE. Columns NW and B report whether the coefficient of interest is statistically significant at 1%, 5%, and 10% levels according to the Newey-West standard errors and the Mammen wild bootstrap quantiles, respectively. We count 44 observations.
*Statistically significant at 10% level.
**Statistically significant at 5% level.
***Statistically significant at 1% level.

TABLE A8 | Determinants of the weights constructed using the Spec. 2 weights.

(a)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.568	0.034	***	***	0.343	0.168	**	***
Dif. 2%	−0.041	0.008	***	***	−0.186	0.034	***	***
EPU	0.00014	0.00016	***	***	−0.002	0.001	**	***
S	0.047	0.005	***	***	0.228	0.022	***	***
Inf. Sur.	−0.022	0.006	***	***	−0.109	0.027	***	***
Glob.	−0.009	0.004	**	***	−0.032	0.018	*	***
Frag.	−0.021	0.011	*	***	−0.094	0.046	**	***
R^2	0.855				0.8634			

(b)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.108	0.329			0.607	0.312	*	
Dif. 2%	−0.044	0.011	***	***	−0.204	0.057	***	***
Dis.	−0.004	0.021			−0.015	0.123		
EPU	−0.00037	0.000129			−0.001	0.001		
S	0.045	0.003	***	***	0.219	0.016	***	***
APP	0.002	0.006			0.003	0.032		
Inf. Sur.	−0.020	0.005	***	***	−0.105	0.027	***	***
Glob.	−0.002	0.008			−0.011	0.030		
Frag.	−0.021	0.009	**	***	−0.096	0.040	**	***
Oil	0.087	0.061			0.614	0.498		
Spread	0.060	0.069			0.280	0.313		
d_2	−0.039	0.019	*	*	−0.187	0.087	**	*
d_3	−0.023	0.027			−0.115	0.120		
d_4	−0.045	0.036			−0.208	0.167		
R^2	0.853				0.8688			

Note: The first column displays the determinants; b_0 is the intercept. LM and Beta report the estimated coefficients using the linear regression model and the beta regression. We report the adjusted R^2 for the LM and adjusted pseudo R^2 for the Beta. We display Newey-West standard errors in column SE. Columns NW and B report whether the coefficient of interest is statistically significant at 1%, 5%, and 10% levels according to the Newey-West standard errors and the Mammen wild bootstrap quantiles, respectively. We count 44 observations.
*Statistically significant at 10% level.
**Statistically significant at 5% level.
***Statistically significant at 1% level.

TABLE A9 | Determinants of the weights constructed using the Spec. 3 weights.

(a)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.669	0.062	***	***	0.881	0.384	**	***
Dif. 2%	−0.056	0.012	***	***	−0.276	0.053	***	***
EPU	−0.001	0.000	**	***	−0.003	0.002	**	***
S	0.052	0.007	***	***	0.304	0.051	***	***
Inf. Sur.	−0.043	0.010	***	***	−0.243	0.036	***	***
Glob.	−0.014	0.007	**	***	−0.036	0.023		***
Frag.	−0.039	0.013	***	***	−0.210	0.062	***	***
R^2	0.816				0.6238			

(b)	LM	SE	NW	B	Beta	SE	NW	B
b_0	0.579	0.431			3.106	3.935		
Dif. 2%	−0.046	0.016	***	***	−0.190	0.107	*	***
Dis.	−0.033	0.030			−0.219	0.214		
EPU	−0.001	0.000	**	**	−0.004	0.002	**	***
S	0.052	0.006	***	***	0.312	0.062	***	***
APP	0.005	0.011			0.026	0.084		
Inf. Sur.	−0.039	0.009	***	***	−0.213	0.056	***	***
Glob.	−0.009	0.010			−0.037	0.050		
Frag.	−0.039	0.008	***	***	−0.201	0.066	***	***
Oil	0.023	0.084			−1.133	2.262		
Spread	0.059	0.081			0.034	0.495		
d_2	−0.064	0.029	**	*	−0.383	0.186	**	
d_3	−0.051	0.031		*	−0.291	0.175	*	
d_4	−0.076	0.045		**	−0.452	0.286		
R^2	0.804				0.5919			

Note: The first column displays the determinants; b_0 is the intercept. LM and Beta report the estimated coefficients using the linear regression model and the beta regression. We report the adjusted R^2 for the LM and adjusted pseudo R^2 for the Beta. We display Newey-West standard errors in column SE. Columns NW and B report whether the coefficient of interest is statistically significant at 1%, 5%, and 10% levels according to the Newey-West standard errors and the Mammen wild bootstrap quantiles, respectively. We count 44 observations.
*Statistically significant at 10% level.
**Statistically significant at 5% level.
***Statistically significant at 1% level.