European Economic Review 178 (2025) 105079

Contents lists available at ScienceDirect

European Economic Review

journal homepage: www.elsevier.com/locate/eer

Density forecasts of inflation: A quantile regression forest approach[☆]

Michele Lenza [a,c,*], Inès Moutachaker [b], Joan Paredes [a]

[a] *European Central Bank, Germany*
[b] *INSEE, France*
[c] *CEPR, United Kingdom*

ARTICLE INFO

JEL classification:
C52
C53
E31
E37

Keywords:
Inflation
Non-linearity
Quantile regression forest

ABSTRACT

Inflation density forecasts are a fundamental input for a medium-term-oriented central bank, such as the European Central Bank (ECB). We demonstrate that a quantile regression forest, which captures general non-linear relationships between euro area inflation (both headline and core) and a broad set of determinants, performs competitively against state-of-the-art linear and non-linear benchmarks and judgmental forecasts. The median forecasts generated by the quantile regression forest exhibit a high degree of collinearity with the Eurosystem inflation point forecasts, displaying similar deviations from "linearity". Given that the Eurosystem's modeling toolbox predominantly relies on linear frameworks, this finding suggests that the expert judgment embedded in the projections may incorporate mild non-linear elements. Finally, we provide a real-time application illustrating how the model is employed to assess risks surrounding the Eurosystem inflation projections in the context of the recent euro area disinflation path.

1. Introduction

The mandate of the European Central Bank (ECB) is to maintain price stability *over the medium-term*. This medium-term orientation implies that sources of fluctuations with temporary effects on inflation are more likely to be "looked through", whereas more persistent inflationary pressures may have a greater influence on monetary policy decisions. Consequently, inflation projections, which synthesize the Eurosystem staff's assessment on inflation dynamics, play a pivotal role in shaping monetary policy decisions. However, economic projections are inherently subject to uncertainty, and monetary policy decisions rely not only on point forecasts but also on a careful evaluation of the likelihood of various hypothetical current and future scenarios, a process defined as "risk assessment".

Modeling inflation dynamics in the euro area remains a significant challenge, given the multitude of potential driving factors, the complexity of their interactions with inflation, and the relatively short historical sample available for econometric analysis (see, for example, Koester et al., 2021). A central debate in the literature concerns whether the relationship between inflation and its determinants is best characterized as linear or non-linear, with important implications for both forecasting accuracy and monetary policy assessment.

The expanding literature on macroeconomics@risk highlights the importance of accounting for non-linearities in the dynamics of key policy-relevant variables, particularly for the robustness of risk assessments conducted by central banks. In particular, Adrian

☆ The authors thank Joscha Beckmann, Laurent Ferrara, Francesco Furlanetto, Domenico Giannone, Philip Lane, Giorgio Primiceri, Isabel Schnabel, Dalibor Stevanovic and several seminar/conference participants for comments. All remaining errors are our own. The views expressed are those of the authors and do not necessarily reflect those of the European Central Bank (ECB), the Eurosystem and the CEPR.

* Corresponding author at: European Central Bank, Germany.
E-mail address: michele.lenza@ecb.europa.eu (M. Lenza).

https://doi.org/10.1016/j.euroecorev.2025.105079
Received 24 May 2024; Received in revised form 27 May 2025; Accepted 27 May 2025
Available online 13 June 2025
0014-2921/

et al. (2019) demonstrate that the dynamics of the upper and lower quantiles of the GDP predictive density are closely linked to fluctuations in comprehensive indices of financial conditions. Moreover, a growing body of research, including Chavleishvili and Manganelli (2019), López-Salido and Loria (2020), Figueres and Jarociński (2020), Adams et al. (2021), Korobilis et al. (2021), Goulet Coulombe et al. (2022), Kiley (2022), Amburgey and McCracken (2023a,b), Botelho et al. (2023), Boyarchenko et al. (2023), Chavleishvili et al. (2023) and Chavleishvili and Kremer (2023) shows that similar considerations extend to the density forecasts of inflation and other macroeconomic variables.

The comprehensive overview of the Eurosystem's economic analysis conducted in Darracq Pariès et al. (2021) as part of the recent ECB strategy review, highlights the predominance of linear models within the Eurosystem's modeling toolkit. For example, Giannone et al. (2014) describes and evaluates in a real-time framework a large BVAR to predict inflation in the euro area. The model has been used to produce a density forecast for HICP headline inflation and its sub-components. In this paper, we introduce a novel non-linear model for euro area inflation density forecasting. Specifically, to capture the relationship between a broad set of potential predictors and inflation, we employ the quantile regression forest (QRF) of Meinshausen (2006), which is a variant of the random forest method developed by Breiman (2001). The random forest is an ensemble technique that aggregates multiple non-linear predictive models, known as regression trees. These regression trees recursively partition the predictor space into (potentially numerous) subsets, referred to as "leaves", and generate predictions by computing the average value of the target variable within each leaf. Quantile regression forests build on the same principles but extend the methodology by estimating the empirical quantiles of the target variable's distribution in the leaves, thereby enabling density forecasting.

Compared to the existing literature, the QRF offers a more flexible approach to capturing non-linear relationships, as it does not impose any specific parametric structure between predictors and the target variables. Unlike many @risk models, which typically assume a linear relationship between predictive quantiles and their determinants, the QRF remains fully data-driven, allowing for more general forms of non-linearity. Moreover, the QRF can seamlessly accommodate a large set of predictors, enabling the inclusion of all potentially relevant information for inflation forecasting. This flexibility represents a significant advantage over conventional @risk applications, which are often constrained by a limited number of explanatory variables. To incorporate a sufficiently comprehensive information set, these models typically rely on summary measures of financial or general economic conditions, whereas the QRF directly utilizes the full range of available data.

We measure headline inflation as the rate of change of the Harmonized Index of Consumer Prices (HICP), and consider an additional measure of "core inflation", specifically the rate of change of the Harmonized Index of Consumer Prices excluding Energy and Food prices (HICPX).[1] The selection of potential determinants for headline and core inflation in our model is broadly motivated by the Phillips Curve framework and include variables capturing inflation expectations, cost pressures, real economic activity and financial conditions.

In our empirical application, we evaluate the QRF density forecasts in a recursive out-of-sample exercise over the 2002–2022 evaluation sample, with a forecast horizon of up to one year ahead. First, we compare the forecast accuracy of the QRF against a state-of-the-art linear benchmark, a combination of a large number of Bayesian VAR models (VARCOMB), as well as two alternative non-linear models: the Generalized Random Forest (GRF) of Athey et al. (2016) and the linear quantile regression (LQR) model, which combines a large number of small-scale linear quantile regression models based on Koenker and Bassett (1978) and Adrian et al. (2019). The comparison with the linear benchmark (VARCOMB) provides insight into whether euro area inflation dynamics exhibit non-linear characteristics. In addition, the comparison with the two non-linear benchmarks allows us to assess: (i) whether the estimation method of the GRF, which is explicitly designed to fit quantiles, improves out-of-sample forecasting accuracy relative to the QRF, and (ii) whether the assumption of linearity in the quantiles embedded in LQR is already sufficient to capture the non-linearity in euro area inflation. To formally evaluate the forecasts, we conduct tests of correct calibration as developed by Rossi and Sekhposyan (2019). We further assess forecast performance and ranking using the continuous ranked probability score (CRPS) of Gneiting and Raftery (2007), a proper scoring rule that jointly evaluates calibration and sharpness. Our findings indicate that the QRF produces well-calibrated density forecasts for euro area inflation and is competitive with state-of-the-art linear and non-linear models, particularly for core inflation.

Among the alternative benchmarks, the linear quantile regression (LQR) exhibits the best overall performance. In particular, for headline inflation, LQR slightly outperforms the QRF and VARCOMB, suggesting that the assumption of linearity in the quantiles is appropriate for euro area headline inflation. In contrast, for core inflation, the QRF demonstrates generally superior performance, indicating that more general non-linear relationships may play a more significant role for core inflation dynamics.

We also find that the relative forecasting performance of the QRF and VARCOMB varies over time. The QRF underperformed during and immediately after the Great Financial Crisis (2007–2009) but outperformed VARCOMB during the extended period of low inflation in the euro area prior to the COVID-19 pandemic. For core inflation, instead, the QRF systematically outperforms VARCOMB across the entire evaluation period. Taken together, the results of our forecasting comparison suggest that euro area inflation dynamics exhibit some degree of non-linearity, with stronger evidence of non-linearity in core inflation compared to headline inflation. Since our measure of core inflation excludes energy and food prices, our results imply that the evidence of non-linearity in energy and food prices is less pronounced than in other inflation components. The stronger presence of non-linearity in core inflation is further highlighted when we analyze the contribution of different predictors to the inflation forecasts using Shapley

[1] In the rest of the paper, we refer to HICPX inflation as to core inflation. Despite the popularity of "exclusion measures" of consumer prices to proxy core inflation, there are numerous other measures of core inflation, with advantages and shortcomings. Policy-makers typically consider a range of different measures rather than settling on a single one. For more details, see for example Lenza (2011) and Ehrmann et al. (2018).

values (see Shapley, 1952; Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017; Buckmann and Joseph, 2022). Our examination of the functional relationships between predictors and inflation shows that most relevant predictors exhibit an essentially linear relationship with inflation. However, inflation expectations display a state-dependent relationship, particularly with core inflation.

We also compare the QRF with both survey-based forecasts (SPF) and institutional forecasts from the published Eurosystem Inflation Projections (BMPE). Since the Eurosystem projections are publicly available only as point forecasts, we restrict this comparison to the median and mean QRF forecasts. For the comparison with the SPF density forecasts, we focus on headline inflation, given the limited availability of SPF core inflation forecasts. Remarkably, the QRF generally outperforms the SPF, particularly at short forecast horizons. In general, QRF density forecasts exhibit greater sharpness around the true value compared to their SPF counterparts. When comparing the QRF median forecasts with the judgmental Eurosystem inflation forecasts, we find that both sets of forecasts exhibit similar accuracy and dynamics, except for headline inflation at short forecast horizons, for which the BMPE forecasts are more accurate. Examining the gaps between our median VARCOMB forecasts and both the median QRF and the Eurosystem forecasts - a rough measure of "distance from linearity" - we observe a strong positive correlation. This suggests that both the QRF and the Eurosystem forecasts exhibit similar deviations from linearity, despite the Eurosystem's reliance on linear modeling frameworks. This finding implies that the judgmental component of the Eurosystem projections incorporates a mild degree of non-linearity in the projected inflation dynamics. Overall, it is remarkable that the QRF produces forecasts that are competitive with those of the SPF and the Eurosystem, especially considering that both the SPF and Eurosystem forecasts incorporate expert judgment informed by news on likely future events, such as VAT changes, fiscal plans or geopolitical developments like the invasion of Ukraine.

Overall, we conclude that the QRF represents a valuable addition to the Eurosystem's forecasting toolbox. However, it should be viewed as a complement rather than a substitute for existing inflation forecasting methodologies, such as linear vector autoregressive models or linear quantile regressions.

At the end of 2022, the model described in this paper was incorporated into regular policy analysis at the ECB. To complement the out-of-sample evaluation, we conduct a real-time assessment of the model's performance from the last quarter of 2022 to the last quarter of 2024, which is the latest quarter available for evaluation at the time of writing. Overall, the real-time performance of the QRF appears very good. Specifically, the model has been employed to assess risks surrounding the Eurosystem projections, and we illustrate how this assessment evolved in real-time during the recent euro area disinflation episode. Notably, the QRF signaled downside risks to the Eurosystem's headline inflation projections, starting in the second half of 2023, which eventually materialized. This highlights the model's potential usefulness in enhancing risk assessments within the Eurosystem's forecasting framework.

In addition to the @risk literature cited above, this paper contributes to the extensive literature on inflation forecasting as well as the broader research on potential non-linearity in inflation dynamics. Regarding the former, comprehensive surveys of the inflation forecasting literature can be found in Faust and Wright (2013) and, more recently, in Banbura et al. (2024). With respect to the latter, a substantial body of research investigates the potential changes in the shape of the Phillips Curve and the underlying factors driving these changes. For a comprehensive survey and systematization of the debate, see Del Negro et al. (2020). Several studies in this literature highlight the differing relationships between inflation and its determinants in high- and low-inflation regimes, or more generally emphasize the state-dependent nature of inflation with its determinants (see, for example Akerlof et al., 1996; Fahr and Smets, 2010; Benigno and Ricci, 2011; Lindé and Trabandt, 2019; Forbes et al., 2021; Clark et al., 2022; Costain et al., 2022; Cavallo et al., 2023; Benigno and Eggertsson, 2023).

In addition, our paper relates to a growing literature on the advantages of ensemble methods, which are increasingly popular in the econometric literature for prediction (Fernandez et al., 2001; Avramov, 2002; Sala-I-Martin et al., 2004; Inoue and Kilian, 2008; Bai and Ng, 2009; Wright, 2009; Rapach and Strauss, 2010; Faust et al., 2013; Ng, 2013; Jin et al., 2014; Varian, 2014; Wager and Athey, 2018; Crump et al., 2018; Clark et al., 2021). Giannone et al. (2021) demonstrate that ensemble methods can be particularly successful due to their ability to effectively handle model uncertainty, while Medeiros et al. (2021) show that the random forest perform well for US inflation prediction. Compared to Medeiros et al. (2021), our study focuses on euro area inflation and, more importantly, on density forecasts, which are central to risk assessment in monetary policy decisions. Despite their critical relevance for policy-making institutions, density forecasts remain relatively underexplored in the machine learning literature.

The remainder of the paper is structured as follows. Section 2 outlines our empirical strategy. Section 3 presents the results of our out-of-sample forecasting accuracy assessment. Section 4 examines the contribution of different predictors to inflation forecasts using Shapley values. Section 5 analyzes the application of the QRF at the ECB for policy analysis and evaluates its real-time performance during the period of disinflation that began in late 2022 in the euro area. Section 6 concludes.

2. Empirical models, data and out-of-sample evaluation

2.1. The quantile regression forest

We adopt a "direct" forecasting scheme, which requires to estimate the relationship between inflation at time t and its determinants at time t–h, for a given forecasting horizon h. Once the model is estimated, we apply it to the most recent data at time t to generate an inflation forecast for t+h. The dependent variable in our model, π_t^h, represents the annualized growth rate of either the Harmonized Index of Consumer Prices (HICP) or of the Harmonized Index of Consumer Prices excluding energy and food prices (HICPX). The price level is denoted generically as P_t below and we produce forecasts for horizons h = 3, 6, 9 and 12 months ahead:

$$\pi_t^h = (12/h) \times [ln(P_t) - ln(P_{t-h})]$$

To formally estimate the non-linear relationship between our target inflation measure, π_t^h, its lags, and a set of determinants x_{t-h}, we specify the following model:

$$\pi_t^h = m(\pi_{t-h}^1 \ldots \pi_{t-h-p}^1; x_{t-h} \ldots x_{t-h-k}) + \varepsilon_t$$

and then obtain an inflation forecast as

$$\hat{\pi}_{t+h}^h = m(\pi_t^1 \ldots \pi_{t-p}^1; x_t \ldots x_{t-k})$$

Rather than imposing a tightly parameterized functional form on $m(.)$, we adopt a more flexible approach that captures general forms of non-linearity by leveraging machine learning techniques. Specifically, we estimate the potentially non-linear relationship between inflation and its determinants using the Quantile Regression Forest (QRF) developed by Meinshausen (2006). This method extends the Random Forest of Breiman (2001) by enabling density forecasting.[2]

A quantile regression forest is an ensemble method that combines the results of multiple non-linear models, specifically regression trees. A regression tree fits a target variable – in this case, headline or core inflation – by recursively partitioning the predictor space into distinct sub-samples. Once the final split is achieved, the predicted value of the target variable within a given sub-sample, defined as "leaf", is typically represented by the sample mean or median for point prediction. In this paper, we focus on density prediction and we follow the approach described in Meinshausen (2006). Rather that providing a single-point forecast, density forecasting is achieved by computing the empirical quantiles of the target variable within each leaf. The partitioning process in a regression tree is conducted through binary recursive partitioning, an iterative procedure that successively splits the data into smaller sub-samples. The process continues until either further splits fail to improve the model according to a statistical criterion, such as the mean squared error (MSE) of the inflation fit, or a predefined stopping rule is reached, ensuring that each leaf contains at least ten observations.

Regression trees are simple models yet they tend to overfit, which makes them bad predicting tools. Many "relatively" uncorrelated regression trees are built to maximize the advantages of combining them, via the following two steps. First, the observations from the original data are bootstrapped with replacement before constructing any new tree. Notice that inflation may be auto-correlated, and we also include two lags of inflation in the inflation determinants, so that the bootstrap procedure does not impair the ability of our model to account for the potential autoregressive dynamics of inflation. Second, the splits are computed, at each node, only by looking at a randomly selected set of the regressors. The default choice for the size of the latter set, which we take in this paper, is to draw a third of the variables for each split. Finally, we set the number of combined regression trees, i.e. the size of the forest, to the default value of 500.[3]

2.2. Benchmark models

We compare the predictions from the quantile regression forest to several benchmark models.

As a state-of-the-art linear model, we consider an equally weighted combination of 500 VAR models, which we define as VARCOMB. This benchmark is chosen to ensure a methodological parallel with the QRF, as both approaches rely on a combination of models, regression trees in the case of the QRF and VAR models in the case of VARCOMB. The main difference between the two lies in the non-linearity captured by QRF, which is absent in VARCOMB. Each individual VAR model includes inflation (headline or core) along with four randomly selected indicators from our dataset. The data are stationarized, before entering the VAR models, and each VAR is estimated with two lags, mirroring the structure of the QRF. The VAR models are estimated using Bayesian techniques. The prior distributions for the lag coefficients and error variances follow the Normal-Inverse Wishart specification and are parameterized to shrink model estimates toward a random walk model, in line with the Minnesota prior (Litterman, 1979; Doan et al., 1984; Banbura et al., 2010).[4] The prior hyperparameters are treated as random variables with their values drawn from the corresponding posterior distribution, following the approach of Giannone et al. (2015).

Our second set of benchmarks includes two alternative non-linear models. The Generalized Random Forest (GRF), proposed by Athey et al. (2016), employs a more tailored rule for estimating conditional quantiles compared to our baseline QRF. Specifically, in the GRF, sample splits for each quantile are determined based on their ability to fit that specific quantile. To ensure comparability, we set the hyperparameters governing the GRF specification to the same values as those used for the QRF.[5] Additionally, we benchmark our results against the Linear Quantile Regression (LQR) model introduced by Koenker and Bassett (1978), which serves as the primary framework in Adrian et al. (2019) and the broader macroeconomics@risk literature. Unlike the QRF, the LQR assumes a linear relationship between the quantiles and the predictors. Precisely, our LQR benchmark consists of a combination of 500 linear quantile regressions, each including two lags of inflation and four randomly selected predictors, mimicking the procedure we follow for VARCOMB.

2 To estimate and process the output of the QRF we use the R package "ranger" publicly available at https://github.com/imbs-hl/ranger.

3 Probst et al. (2019) discusses the default specification choices for random forests and quantile regression forests and also elaborates on the techniques to tune the model.

4 The data are stationary, so we center the prior on all the lag coefficients to zero.

5 To derive the GRF results, we used the R package "grf" publicly available at https://grf-labs.github.io/grf/.

Our third benchmark consists of headline inflation density forecasts from the ECB Survey of Professional Forecasters (SPF).[6] The SPF is conducted on a quarterly basis, with participants comprising experts affiliated with financial or non-financial institutions across Europe. For this analysis, we collected historical vintages of SPF headline inflation forecasts, aggregated across experts, covering the period from February 2002 until November 2022. The density forecasts are available for two definitions of inflation, (i) year-on-year inflation[7] and (ii) inflation within the current calendar year.[8] In terms of methodology, SPF participants use a mix of econometric models and expert judgment to generate their forecasts. They provide a probabilistic assessment of inflation outcomes by assigning probabilities to predefined ranges of inflation values.

Our final benchmark consists of the Eurosystem headline and core inflation projections (BMPE, in short).[9] These institutional forecasts are produced quarterly and published at the beginning of the third month of each quarter. For our analysis, we consider BMPE vintages from March 2002 to December 2022 for both headline and core inflation. Since the BMPE does not provide a density forecast for a significant portion of the sample, we restrict our comparison to the median and mean QRF forecasts. Methodologically, BMPE projections are based on a combination of model-based analysis and expert judgment.

Additionally, in Appendix B, we provide a comparison of the median QRF forecasts with those from a random walk model (RW), a widely used benchmark for non-forecastability. Following the approach of Atkeson and Ohanian (2001), the RW model forecasts inflation at time "t+h" as:

$$\hat{\pi}^{12}_{t+h} = \pi^{12}_{t}$$

2.3. Data

In addition to headline HICP and HICP excluding energy and food – our two target variables – our database includes 59 additional variables, as listed in Appendix A. Furthermore, we always include two lags of inflation among the predictors, bringing the total cross-section of predictors to 61. The dataset is sourced from the ECB Statistical Data Warehouse (SDW) and draws on a variety of original sources. Broadly speaking, the selection of variables is inspired by the Phillips Curve framework, covering key macroeconomic and financial dimensions, and follows a similar approach to de Bondt et al. (2018).

Specifically, we include measures of cost pressures (for example, commodity prices, exchange rates, wages and producer prices); survey and hard data on economic activity (for example, European Commission surveys employment expectations, confidence measures, industrial production, euro area business cycle indicators, various productivity measures); measures of inflation expectations (for example, European Commission surveys on prices); and financial variables (for example, interest rates, monetary aggregates, asset prices, bank lending).

Our sample ranges from December 1991 to December 2022 and the frequency of the data is monthly. We stationarize the data, when needed. We also de-seasonalize the data in accordance with our out-of-sample logic. Specifically, for all vintages of our out-of-sample exercise, we estimate the seasonal components by using only the data which would have been available to a forecaster in that vintage. See Appendix A for more details.

2.4. Out-of-sample evaluation

All our out-of-sample exercises follow a recursive updating scheme. Beginning with the comparison among QRF, GRF, VARCOMB, and LQR – all of which can be re-estimated at a monthly frequency – we proceed as follows. First, we estimate our models with data up to December 2001 (which is our first "t") and we produce forecasts for inflation at the three, six, nine and twelve months horizon (t+h). Then, we continue to update the estimation sample by adding one month of data at a time, and we repeat all the steps of the forecasting exercise until exhaustion of the sample. Our evaluation sample ranges until December 2022.

For the comparison with the SPF and BMPE, we adjust our recursive updating scheme to align with the publication dates of these survey-based and institutional forecasts, which are issued at a quarterly frequency. Specifically, for this comparison, we run the QRF and VARCOMB forecasts only once per quarter, ensuring that the data availability at the time of forecasting is comparable to that of the SPF and BMPE forecasters. This adjustment ensures a consistent evaluation framework, minimizing differences in information sets across forecasting methods. Further details on how we match the timing of forecasts across the different models are provided in Appendix A.[10]

The target variable used to assess forecasting accuracy for both headline and core HICP is expressed in year-on-year growth rates. We adopt this convention because some of the benchmark forecasts used for comparison – such as the BMPE – are not seasonally adjusted.

[6] Details on the survey and the historical data are available at https://www.ecb.europa.eu/stats/ecb_surveys/survey_of_professional_forecasters/html/index.en.html. We exclude SPF core inflation denisty forecasts from the evaluation because they were collected only starting in 2017 and the sample is too short for the evaluation.

[7] For example, for the vintage in the first quarter of the year "t", the experts provide an assessment of inflation between the fourth quarter of year "t–1" and the fourth quarter of year "t".

[8] The concept of inflation in the current year "t" is, effectively, the average year-on-year growth rate of inflation over the four quarters of year "t".

[9] See https://www.ecb.europa.eu/pub/projections/html/index.en.html for more information on the BMPE projections.

[10] Notice, however, that as our data is ex post revised, we are not able to reproduce the exact same data releases forecasters would have in real-time. In Section 5 we present a fully real-time analysis of the QRF forecasts for the period 2022Q4–2024Q4.

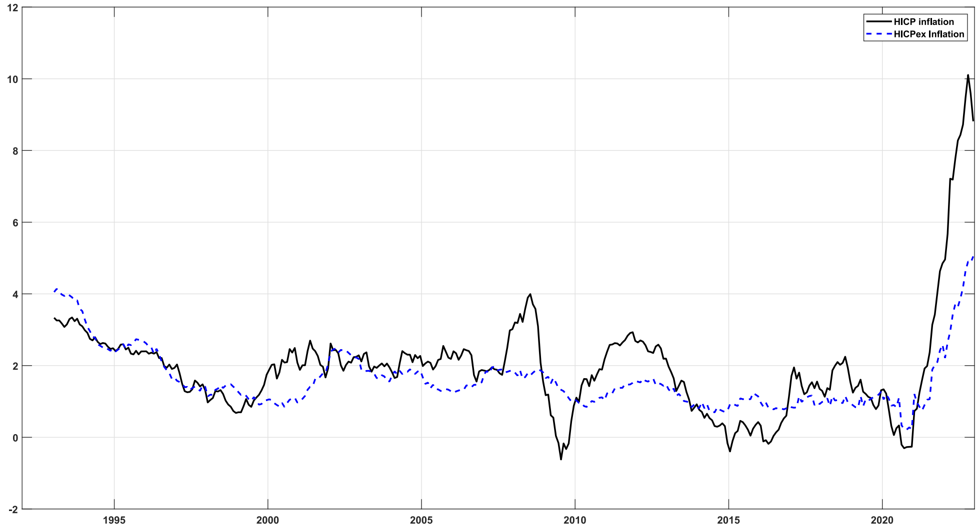

Fig. 1. Headline and core inflation.
Note: Headline inflation: black solid line; Core inflation: blue dashed line. Inflation is defined in terms of year-on-year growth rates of prices and the sample ranges from January 1993 to December 2022.

In other words, for a given forecast horizon h, we evaluate forecasting accuracy using the following transformation of the target variable[11]:

$$\pi_{t+h}^{target} = ln(P_{t+h}) - ln(P_{t+h-12})$$

The only exception to this approach applies to the comparison of density forecasts with the SPF for the *current year*. In this case, we align with the SPF's reporting convention, which expresses current-year inflation as the average year-on-year growth rate over the four quarters of the year.

To assess the ability of different models to capture inflation dynamics across different regimes, we evaluate both the average forecasting accuracy over the entire sample and the evolution of the forecasting accuracy over time (Giacomini and Rossi, 2009, 2010; Rossi and Sekhposyan, 2016).

3. Results

Fig. 1 presents the year-on-year growth rates of the HICP (solid black line) and HICPX (dashed blue line)—the two target measures of headline and core inflation used in our analysis.

The figure illustrates the different inflation regimes that the euro area has experienced over time. Notably, after converging to around 2% in the early 1990s, both headline and core inflation remained relatively stable until the onset of the 2007–2009 financial crisis. In the years leading up to the financial crisis, headline inflation rose sharply, primarily driven by a significant increase in commodity prices, while core inflation remained stable. The recession following the crisis triggered a sudden and pronounced decline in headline inflation, whereas core inflation declined more gradually. Following the Great Recession and an initial inflation rebound, the euro area entered a prolonged period of relatively low inflation. In the post-pandemic period, both headline and core inflation surged to historically unprecedented levels within the euro area sample. Kuik et al. (2022) analyze the role of energy market disruptions – particularly those triggered by the Russian invasion of Ukraine – in fueling the euro area inflation surge. Initially, inflationary pressures were concentrated in the energy component, but as 2022 progressed, price increases became more broad-based, eventually impacting the entire consumption basket and leading to a sharp rise in core inflation as well.[12]

3.1. Evaluation of the density forecasts

3.1.1. Comparison with VARCOMB, GRF and LQR

Fig. 2 presents the 16th to 84th predictive quantile ranges for headline inflation, from the QRF (Panel a) and VARCOMB (Panel b), as obtained from the recursive out-of-sample exercise described in the previous section. To better illustrate changes in the width and

[11] It may be worth reminding here that, as described above, to produce a forecast for "t+h" the variable we fit in our models is instead the annualized growth rate of prices between "t–h" and "t".

[12] See Giannone et al. (2014) for a quantification of the pass-through of commodity price shocks to core inflation components. See Giannone and Primiceri (2024) for a critical discussion on the prevailing narrative for the recent surge in euro area inflation and the subsequent disinflation path.

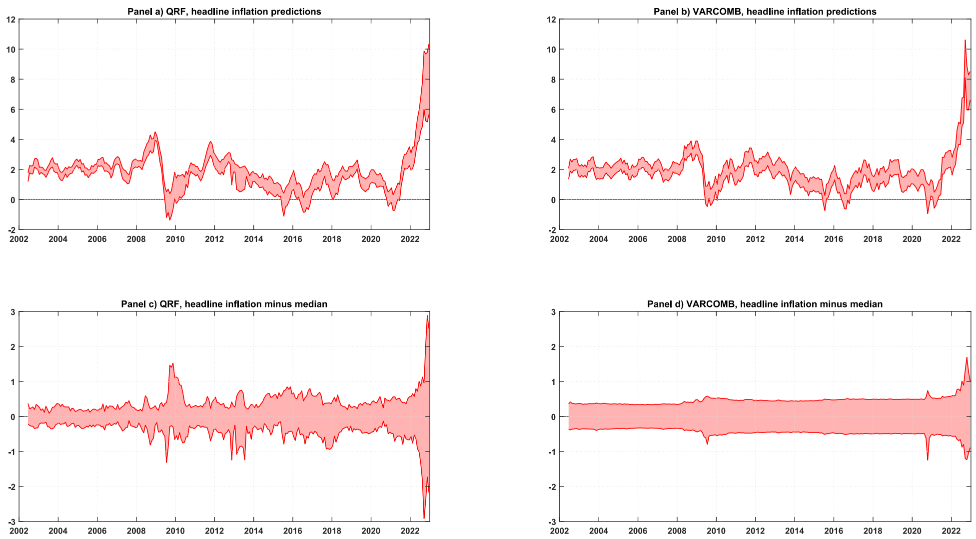

Fig. 2. Headline inflation, prediction densities and ranges, QRF and VARCOMB, h = 6.
Note: Left panels: QRF; Right panels: VARCOMB. Top panels: h = 6 predictive density of year-on-year headline inflation, 16th to 84th quantiles; Bottom panels: 16th to 84th predictive range obtained by subtracting the median forecasts from the quantiles. Sample: June 2002–December 2022.

shape of the predictive intervals, the bottom panels display the 16th and 84th quantiles after subtracting the median forecasts. This transformation allows for a more precise assessment of how forecast uncertainty evolves over time. We focus on the six-month-ahead horizon for headline inflation; however, our conclusions remain robust across different inflation measures (headline or core) and forecasting horizons.

The VARCOMB predictive ranges exhibit gradual changes, driven primarily by the recursive nature of the estimation process and by episodes of significant volatility in the underlying variables. In contrast, the QRF predictive ranges display more dynamic variations in width and shape, reflecting a higher degree of adaptability to changing economic conditions. A key question that arises is whether the greater volatility in the QRF quantiles translates into improved out-of-sample predictive accuracy.

As a first step in our forecast evaluation, we assess whether the density forecasts produced by the QRF, VARCOMB, GRF, and LQR are correctly calibrated. A density forecast is considered well-calibrated if, when it "assigns a certain probability to an event, then the event should occur with the stated probability over successive observations" (Elliott and Timmermann, 2016).

Defining $p(y_t)$ as a generic density forecast, we assess its calibration accuracy by testing whether the Probability Integral Transform (PIT)—i.e., the cumulative density function (CDF) corresponding to $p(y_t)$- of the realized values of y_t follows a uniform distribution $U(0,1)$ (Diebold et al., 1998). Several methodologies have been proposed in the literature to evaluate density forecast calibration (see, for example Diebold et al., 1998; Berkowitz, 2001; Corradi and Swanson, 2006; Hong et al., 2007; González-Rivera and Sun, 2015; Knüppel, 2015). In this study, we adopt the test procedure developed by Rossi and Sekhposyan (2019), which is based on the Kolmogorov–Smirnov test and also provides a graphical representation of calibration results.[13] Fig. 3 presents the calibration results for headline inflation, while Fig. 4 reports the corresponding results for core inflation.[14]

For headline inflation, we generally accept the null hypothesis, indicating that the density forecasts from all models are well calibrated. For core inflation, however, the results present a slightly different picture. All models exhibit some degree of miscalibration, with the issue being particularly pronounced for VARCOMB and LQR. Specifically, for these two models, we reject the null hypothesis of well-calibrated forecasts for forecasting horizons beyond six months.

While calibration is a desirable property for density forecasts, it is not sufficient on its own. Hamill (2000) emphasizes that calibration is merely a necessary condition for an ideal forecaster—one that perfectly captures the true cumulative distribution function. Gneiting et al. (2007) further argue that, given correct calibration, maximizing sharpness improves the approximation of the ideal forecaster. For this reason, in addition to calibration, we evaluate the relative accuracy of the density forecasts using a proper scoring rule, namely the Continuous Ranked Probability Score (CRPS) proposed by Gneiting and Raftery (2007). The CRPS

[13] For the implementation of such test procedure, we need the draws from the predictive density of the different models. Hence, for the sake of running this test procedure, we follow the practice in the literature (see, for example, Adrian et al., 2019) and fit a skew-t distribution (Azzalini and Capitanio, 2003; Adrian et al., 2019). See the details in Appendix A. Strictly speaking, we would not need to fit a distribution to the VARCOMB forecasts, which are already a draw from the posterior distribution of VARCOMB. We fit the skew-t distribution also for VARCOMB only for comparability purposes, but all the results in the paper are robust to using the original posterior draws. As also shown in Montes Galdon et al. (2022), the skew-t distribution is an appropriate choice because it is a flexible parametric density that allows for fat tails, as well as asymmetries.

[14] Notice that our forecasts are multi-step, since we look at forecast horizons ranging from three to twelve months ahead. Hence, for our tests we follow the suggestion of Rossi and Sekhposyan (2019) and we compute critical values from a block version of the weighted bootstrap of Inoue (2001). The computations are carried out by using the replication codes kindly provided in Rossi and Sekhposyan (2019).

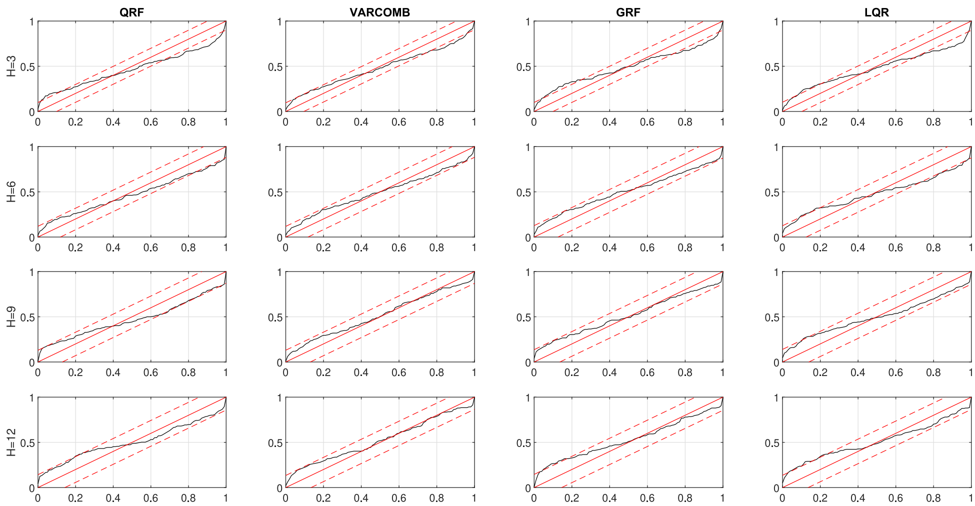

Fig. 3. Headline inflation, test of uniformity of PITs.
Note: Red lines: 1% critical values of the Kolmogorov–Smirnov test of PIT uniformity (dashed) and 45% degree line; Black line: Cumulative distribution function (CDF) of the PITs. If the probability density function of the PIT is a U(0,1), the CDF should be the 45% degree line. The four rows correspond to the four forecasting horizons in the paper and the four columns, from left to right, to the QRF, VARCOMB, GRF and LQR models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

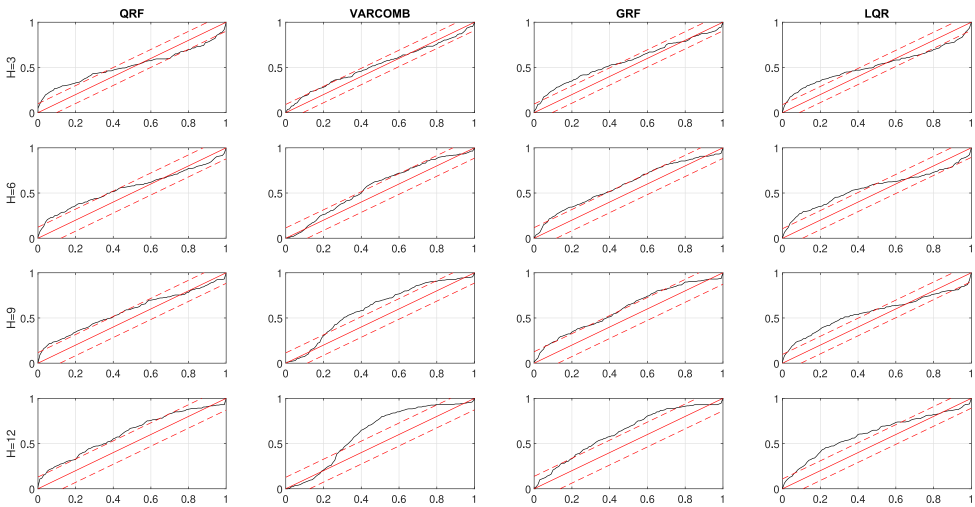

Fig. 4. Core inflation, test of uniformity of PITs.
Note: Red lines: 1% critical values of the Kolmogorov–Smirnov test of PIT uniformity (dashed) and 45% degree line; Black line: Cumulative distribution function (CDF) of the PITs. If the probability density function of the PIT is a U(0,1), the CDF should be the 45% degree line. The four rows correspond to the four forecasting horizons in the paper and the four columns, from left to right, to the QRF, VARCOMB, GRF and LQR models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

quantifies the "distance" between the predictive cumulative distribution function and the empirical cumulative distribution function associated with the observed realizations of the target variable. A lower CRPS indicates a more accurate density forecast.[15] Scoring rules such as the CRPS simultaneously assess calibration and sharpness (i.e., the concentration of density forecasts). Thus, examining the CRPS allows us to complement the calibration assessment conducted earlier. Another key advantage of scoring rules is that they enable us to rank different models reports the CRPS results for headline inflation (Panel A) and core inflation (Panel B), expressed as ratios relative to the CRPS of the QRF. A value greater than one indicates that the QRF outperforms the corresponding model in terms of density forecasting accuracy (see Table 1).

[15] We compute the CRPS directly from the quantiles of our predictive distributions. In the working paper version Lenza et al. (2023), we fit a skew-t distribution for all the models and compute the CRPS from the draws of that distribution and the results remain the same.

Table 1
CRPS of different models for headline and core inflation.

Horizon	VARCOMB	GRF	LQR
Panel a: Headline inflation			
h = 3	1.05	1.03	0.93
h = 6	1.03	1.02	0.92
h = 9	1.03	1.02	0.94
h = 12	1.03	1.05	0.97
Panel b: Core inflation			
h = 3	1.47	1.13	1.06
h = 6	1.42	1.17	1.06
h = 9	1.38	1.14	1.01
h = 12	1.36	1.15	1.00

Note: Relative CRPS for VARCOMB (second column), GRF (third column) and LQR (fourth column). A value larger than one indicates for a specific model and horizon indicates that the QRF outperforms that model at that horizon.

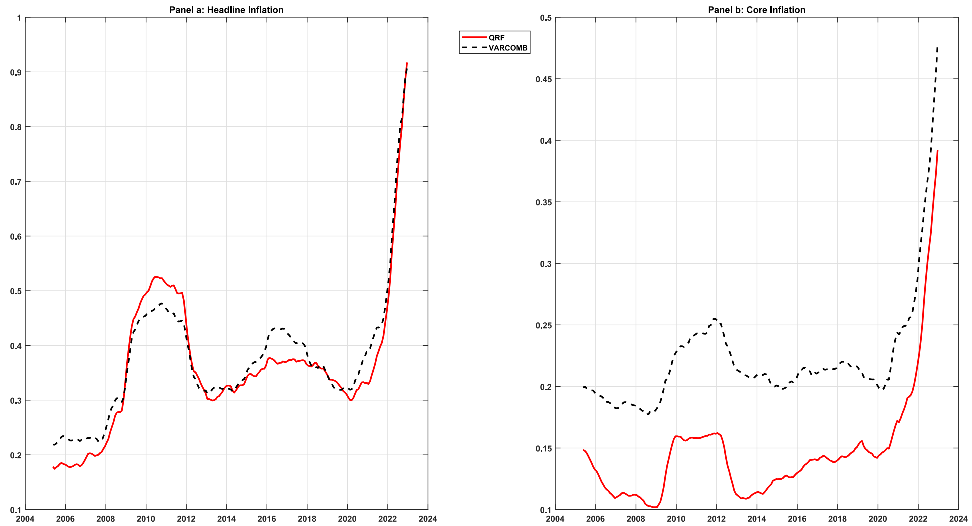

Fig. 5. CRPS, three years rolling window, headline and core inflation, h = 6.
Note: Red solid line: QRF; Black dashed line: VARCOMB. The value on the vertical axis at each point refers average CRPS over a rolling window of 36 months.

For headline inflation, we find that all models exhibit similar forecasting accuracy across all horizons, with the LQR model slightly outperforming the others. In contrast, for core inflation, the QRF either matches or outperforms other models across all forecast horizons. Overall, our results indicate that the QRF is competitive with state-of-the-art linear and non-linear benchmarks, particularly at short horizons. Interestingly, the GRF does not improve upon the QRF, suggesting that an estimation method more explicitly tailored to quantile fitting does not provide significant advantages in this empirical application.

These findings lead us to conclude that the QRF represents a valuable addition to the Eurosystem's inflation forecasting toolbox, enhancing its ability to assess inflation risks and uncertainty. Given its comparable accuracy to state-of-the-art linear models for headline inflation, the QRF should be viewed as a complement rather than a substitute for the Eurosystem's predominantly linear forecasting toolbox.

To evaluate the QRF's ability to capture inflation dynamics across different regimes (Giacomini and Rossi, 2009, 2010; Rossi and Sekhposyan, 2016), Fig. 5 presents the CRPS of QRF and VARCOMB, computed over rolling three-year windows. For brevity, we focus on the six-month-ahead forecasting horizon, reporting results separately for headline inflation (left panel) and core inflation (right panel). This rolling-window analysis provides insights into the time variation in forecast performance, helping to identify periods where non-linearity may play a more significant role in inflation dynamics.

Focusing first on headline inflation, we find that VARCOMB outperforms QRF in capturing the rapid inflation rebound following the Great Recession. Specifically, VARCOMB detects the inflation trough earlier and exhibits less reactivity than the QRF throughout the crisis period. This result aligns with the findings of Ferrara et al. (2015) and Bobeica and Jarociński (2019), which suggest that linear models – when accounting for large shocks – can accurately describe inflation dynamics around the Great Recession. However, the QRF adapts much faster than VARCOMB to the prolonged period of low inflation that characterized the pre-COVID decade. The improved accuracy of the non-linear model in this episode suggests that low-inflation regimes may be associated with different inflation dynamics than high-inflation regimes, as proposed in Forbes et al. (2021). Nevertheless, given the relatively short

Table 2
Relative CRPS of SPF relative to QRF.

Horizon	SPF
Year-on-year, h = 12	0.94
Current year, Q1	1.07
Current year, Q2	1.22
Current year, Q3	1.45
Current year, Q4	2.63

Note: A value larger than one indicates that the QRF is more accurate than the alternative model.

evaluation sample, distinguishing between high- and low-inflation regimes remains challenging. In the most recent period, both the QRF and VARCOMB struggled to fully capture the high-inflation regime. The right panel of Fig. 5 presents a contrasting picture for core inflation: unlike for headline inflation, the QRF consistently outperforms VARCOMB across the sample, highlighting its superior predictive accuracy for core inflation dynamics.

Taking all results together, our findings suggest that non-linearity in inflation dynamics cannot be ruled out for core inflation. In contrast, for headline inflation, the linear VARCOMB benchmark remains competitive with non-linear alternatives. The key distinction between headline and core inflation is that core inflation excludes energy and food prices from the HICP, two components largely driven by global commodity price dynamics. This suggests that the pass-through of commodity prices to headline inflation – particularly via energy prices – is characterized by linear dynamics. When commodity price volatility dominates headline inflation dynamics, the underlying non-linearity in core inflation remains less apparent. In such periods, linear models remain competitive for forecasting headline inflation, as the inflation process is primarily shaped by external cost-push shocks rather than intrinsic non-linearities in domestic inflation dynamics.

3.2. Comparison with survey and institutional forecasts

3.2.1. Comparison with the Survey of Professional Forecasters (SPF)

The SPF provides forecasts for year-on-year headline inflation only at the one-year-ahead horizon. However, inflation forecasts for the current calendar year, which approximate the average year-on-year inflation over the calendar year, are reported in the SPF at each quarter of the year. This feature allows us to assess the SPF's density forecasting accuracy at shorter than one year horizons as well.

Table 2 presents the relative Continuous Ranked Probability Score (CRPS) of the SPF forecasts compared to the QRF for (i) year-on-year inflation forecasts at the one-year-ahead horizon, and (ii) current-year headline inflation forecasts, evaluated in each of the four quarters of the year.

The results clearly indicate that the QRF outperforms the SPF for short-horizon forecasts. While this comparison has some limitations – since the SPF forecasts are conducted in real-time, whereas the QRF is evaluated using ex post revised data – we find the results nonetheless remarkable because the SPF is a judgmental forecast. Hence, the SPF can incorporate valuable forward-looking information, such as expected policy changes or geopolitical developments, which may not be embedded in the QRF's information set.

3.2.2. Comparison of point forecasts with BMPE

The Eurosystem inflation forecasts (BMPE) have historically been reported as point forecasts for a significant portion of the sample under analysis. Consequently, we limit our comparison to point forecast accuracy. The baseline QRF point forecast is computed as the median of the inflation values in the leaves. Table 3 presents the relative root mean squared errors (RMSE) for the BMPE, VARCOMB, and an alternative QRF point forecast, for which the QRF point forecast is computed by taking the mean rather than the median of the observations in the leaves of the regression trees. The relative RMSE is calculated as the ratio of each model's RMSE to the RMSE of our baseline QRF point forecast. A value greater than one indicates that the median QRF forecast is more accurate than the alternative forecast.

The QRF point forecasts are generally comparable in accuracy to the BMPE forecasts, except at short horizons for headline inflation. This result is particularly noteworthy, given that BMPE forecasts are produced through a highly refined and sophisticated analytical process by Eurosystem forecasters. Additionally, BMPE forecasts incorporate expert judgment, allowing them to account for expected economic developments that may not be fully captured by the variables used in the QRF model. Consistent with previous findings on density forecasts, the VARCOMB model is less accurate than QRF for core inflation, while its accuracy for headline inflation is comparable to that of the QRF. However, VARCOMB's relative performance in point forecasts appears to be better than in density forecasts, suggesting that non-linearity plays a more significant role in density forecasting than in point forecasting. These results remain robust when using the mean instead of the median to compute QRF point forecasts, as the RMSEs of the QRF-mean forecasts are generally very close to those obtained using the median QRF forecasts.

Fig. 6 presents the headline QRF inflation forecasts alongside the BMPE and observed inflation at the two-quarter-ahead forecast horizon.

Even after accounting for the fact that the similarity between the two sets of forecasts is amplified by the use of year-on-year growth rates, the BMPE forecasts exhibit a high degree of collinearity with the QRF forecasts. Given that the Eurosystem forecasts

Table 3
Relative RMSE for headline and core inflation.

Horizon	BMPE	VARCOMB	QRF-mean
Panel a: Headline inflation			
h = 1Q	0.79	1.04	1.00
h = 2Q	0.89	1.05	0.98
h = 3Q	0.95	0.93	0.97
h = 4Q	0.98	0.89	1.00
Panel b: Core inflation			
h = 1Q	1.06	1.42	1.02
h = 2Q	1.03	1.41	1.07
h = 3Q	1.02	1.05	1.02
h = 4Q	0.94	0.88	1.01

Note: relative RMSE. A number larger than one indicates that the median QRF forecast is more accurate than the forecasts from a specific model (column) and a specific horizon (row).

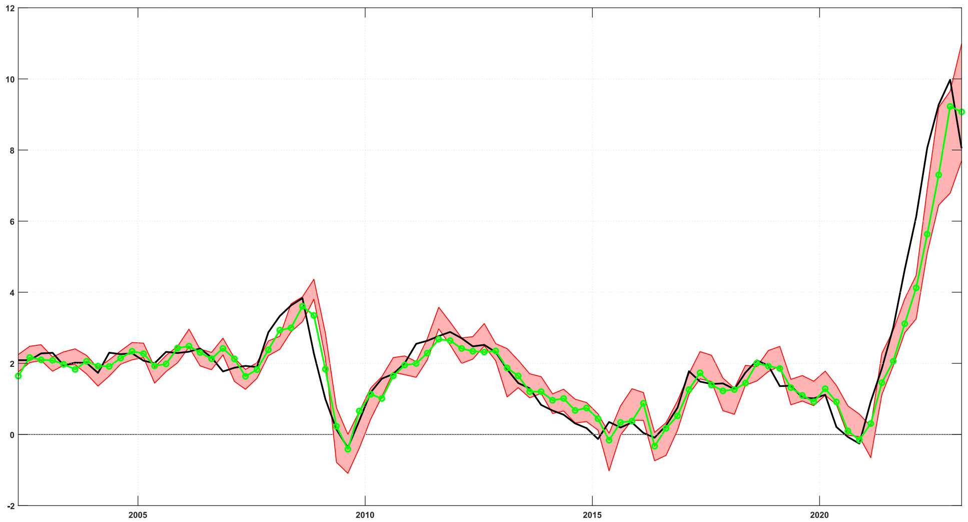

Fig. 6. Headline inflation, density forecasts of QRF and BMPE, 2 quarters ahead.
Note: Black solid line: year-on-year growth rate of HICP (headline inflation); Red area: 16th to 84th quantiles of the QRF density forecasts for the horizon of six months ahead, year on year growth rate of HICP; Green line with circles: BMPE projections for the horizon of six months ahead, year on year growth rate of HICP.

are produced using a predominantly linear modeling framework, this result suggests that the judgmental component embedded in the BMPE projections introduces a degree of non-linearity.

Fig. 7 illustrates the differences (gaps) between the median QRF forecast and the median VARCOMB forecast, as well as the BMPE forecast versus VARCOMB, providing a rough measure of the "distance from linearity" in both forecasts.

The two gaps are strongly correlated, with correlation coefficients ranging between 0.4 and 0.7 across the four forecasting horizons analyzed. This finding suggests that judgment systematically introduces some element of non-linearity into the Eurosystem projections.

4. Shapley values: interpretation of the forecasts and their functional forms

In this section, we analyze the drivers of our QRF predictions. To do so, we leverage recent advances in the machine learning literature (see Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017; Buckmann and Joseph, 2022), which recommend using the concept of Shapley values (Shapley, 1952) to quantify the contributions of different variables to the forecasts. In essence, the Shapley value of a specific variable for a given forecast represents its average marginal contribution across all possible coalitions of variables—i.e., all possible combinations of predictors included in the model.[16] The marginal contribution of a variable to a coalition is defined as the incremental impact of adding that variable to the coalition, after integrating out the effects of all variables not

[16] In the literature on machine learning, the variables used as predictors are also defined as "features".

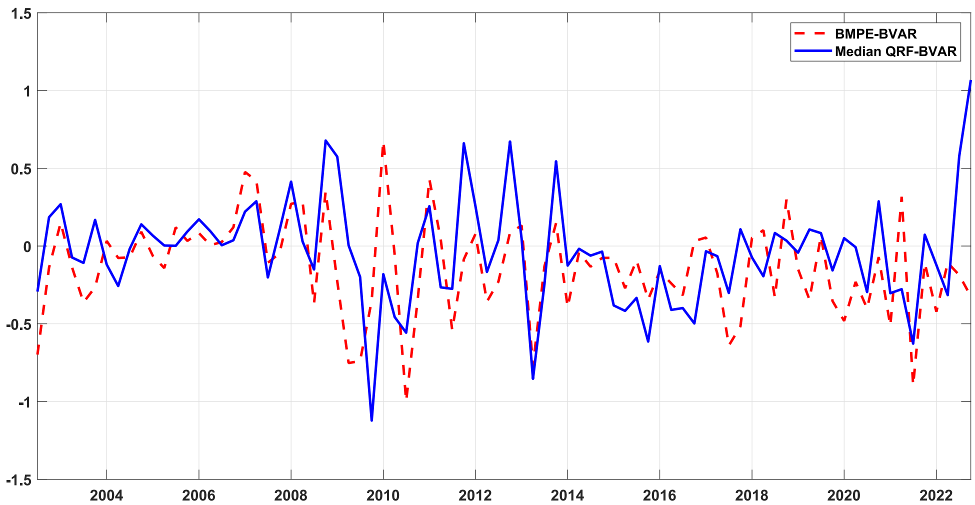

Fig. 7. Gaps BMPE and QRF (median) versus linear BVAR.
Note: Solid blue line: six months ahead (median) QRF forecast of headline inflation minus corresponding VARCOMB forecast; Dashed red line: six months ahead BMPE forecast of headline inflation minus corresponding VARCOMB forecast.

included in the coalition. In the special case of a linear regression where all predictors are orthogonal, the Shapley value of a variable simplifies to the product of the regression coefficient and the deviation of that variable from its mean.[17]

Generally, estimating Shapley values in the presence of a large set of potentially correlated predictors is inherently complex. This complexity arises from two main challenges: the large number of possible coalitions that must be considered in the computation and the difficulty of accurately modeling cross-correlations among predictors when integrating out the effects of "off-coalition" variables. To address these challenges, we adopt the method developed by Lundberg et al. (2019), which leverages the structure of regression trees underlying the QRF to accelerate the computation of Shapley values and effectively handle the issue of correlated predictors.[18]

We employ Shapley values to achieve two key objectives. First, we identify the predictors that drive the inflation outlook. Given that the predictors in our dataset exhibit substantial correlation, as is typically the case for macroeconomic and financial variables, our analysis does not provide a fully causal narrative of inflation dynamics. Instead, the Shapley value decomposition allows us to determine which variables contribute the most to the model's inflation forecasts, thereby highlighting the sources from which the QRF extracts signals to generate its predictions. As an illustration, Fig. 8 presents the results of the Shapley value analysis for headline inflation (left panel) and core inflation (right panel) at the six-month-ahead forecasting horizon over the 2019–2022 period. To facilitate interpretation, individual variables have been classified into groups (see Appendix A), and their Shapley values have been aggregated to compute the overall contribution of each variable group to the predictions.[19]

Between 2019 and 2022, we can distinguish two distinct phases in the QRF's inflation forecasting dynamics. In the first phase, before the recent inflation surge, the QRF primarily relied on financial variables, particularly interest rates, to indicate that inflation was likely to remain low. This is consistent with the historically low levels of interest rates observed during that period. In the second phase, as the QRF began to forecast rising inflation, it extracted the strongest signals from inflation expectations, producer price indices, and real activity indicators. Other groups of variables made negligible contributions to the forecasts. Notably, commodity prices did not play a major role in predicting headline or core inflation. This can be attributed to the fact that other indicators – such as inflation expectations – may have already incorporated the signal that rising commodity prices would translate into higher inflation.

The groups of variables highlighted above have also been among the most significant contributors to the QRF predictions over the full evaluation sample. Table 4 presents the top seven selected variables that contributed the most to headline and core inflation forecasts at the six-month-ahead horizon.[20] The ranking is based on the mean absolute value of each variable's contribution over the out-of-sample evaluation period.

As shown in the table, short-term interest rates, inflation expectations, real activity indicators, and producer prices emerge as key predictors of inflation.

The second key objective of Shapley values is to investigate the functional forms captured by the QRF model. Specifically, we seek to determine which functional relationships best characterize the link between the predictors and headline/core inflation.

[17] See Aas et al. (2020) for a derivation of this result.
[18] In practice, we carry out the computation of Shapley values by using the Tree SHAP package of Lundberg et al. (2019).
[19] The Shapley values of all the variables in the predictor set sum up to the deviation of the predicted value from the average value of the target variable.
[20] Seven is roughly 10% of the variables in our set of predictors.

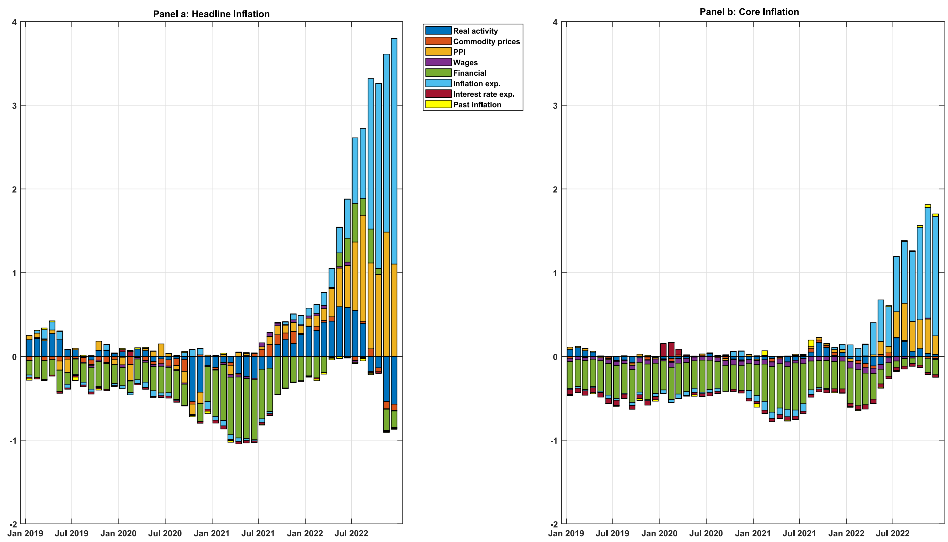

Fig. 8. Decomposition of the h = 6 QRF median forecast.
Note: Shapley values associated with different groups of predictors, sample 2019–2022, horizon of six months ahead. Left panel: Headline inflation; Right panel: Core inflation.

Table 4
Top contributors to six-months ahead forecast.

	Headline inflation	Core inflation
1	Euribor 3-Months	Euribor 3-Months
2	Building permits	Ten-Year Govt Bond Yield
3	Industry survey - selling price expectations for the 3 months ahead	Consumer survey - price trends over next 12 month
4	Consumer survey - price trends over next 12 month	Long-term interest rate future (6 months, DE)
5	Unemployment rate	Unemployment rate
6	Ten-Year Govt Bond Yield	Inflation rate future (6 months, DE)
7	PPI, domestic sales, Total Industry	Indicator of negotiated wage rates - total excluding bonuses

Note: Ranking of top contributors in terms of absolute mean of Shapley value over the evaluation sample, six months ahead horizon. Left column: Headline inflation; Right column: Core inflation.

Fig. 9 presents the Shapley values of a selected set of variables for headline (red circles) and core inflation (blue circles), plotted against the historical values of these variables. The selected variables are among the most relevant inflation indicators in the full sample, as reported in Table 4. These scatter plots provide a rough indication of the type of relationship that the QRF estimates between each predictor and inflation for the six-month forecasting horizon.[21] If the contributions of the predictors (vertical axis) as a function of their values (horizontal axis) lie on a relatively straight line, we can infer that the relationship captured by the QRF is approximately linear. Conversely, deviations from linearity in the scatter plots suggest that the QRF captures non-linear dynamics in the inflation-predictor relationships.

For three of these variables, the relationship captured by the QRF appears to be approximately linear. For instance, the relationship between the Euribor and inflation is essentially linear and positive, as expected: high short-term interest rates signal that inflation is expected to be elevated in the future, necessitating "leaning against the wind" by the central bank. The only notable non-linearity in the Euribor-inflation relationship arises when the ECB reaches the effective lower bound on interest rates. Similarly, the panel on the unemployment rate suggests the presence of a Phillips Curve-type correlation, with a mild negative slope. The pass-through from wages to inflation also appears to be roughly linear. However, for the inflation expectations measure reported in the chart, "price trends expected over the next twelve months", the relationship with headline and, particularly, core inflation, is clearly non-linear. This measure is defined in terms of the balance of survey respondents, where a positive number indicates that there is a larger share of respondents expect prices to increase rather than decrease. The figure suggests that when the share of respondents expecting higher inflation surpasses a certain threshold, inflation tends to accelerate markedly. The fact that the non-linearity is more pronounced for core inflation aligns with our previous finding that the QRF outperforms the linear VARCOMB benchmark in predicting core inflation, reinforcing the conclusion that non-linearities play a greater role in core inflation dynamics than in headline inflation.

[21] Notice that, differently from the rest of the paper, here we are estimating the model for inflation six month ahead on the full sample and, hence, Fig. 9 reports the in-sample Shapley values.

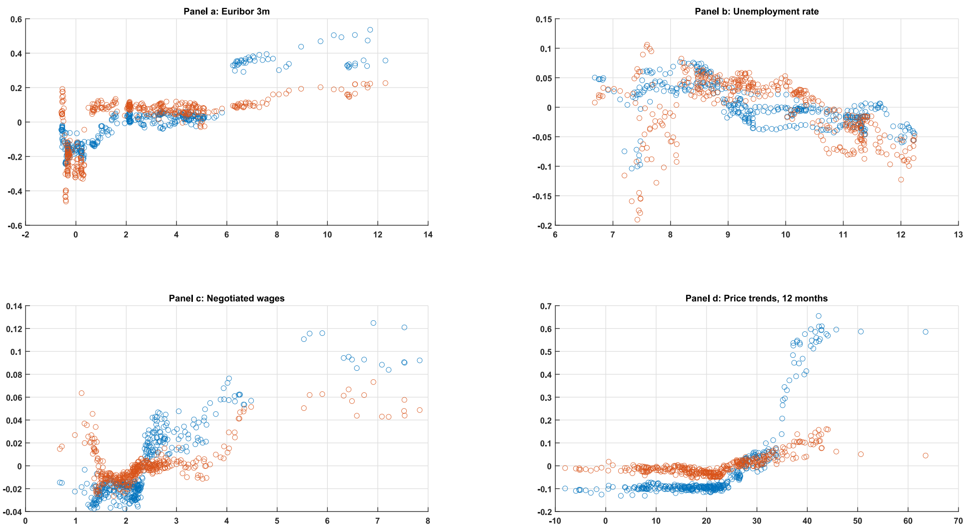

Fig. 9. Top contributors in terms of Shapley values.
Note: Vertical axis: in-sample Shapley values for the variable indicated in the title for headline inflation (red) and core inflation (blue). Horizontal axis: value of the variable indicated in the title. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5. The QRF performance in real-time: the case of the 2023–2024 euro area disinflation

In the last quarter of 2022, the QRF was integrated into the ECB's modeling toolbox to support the preparation of short-term inflation projections. The model serves as an additional tool to assess risks surrounding the official Eurosystem projections. An example of its application can be found in Lane (2024).[22]

Since the fourth quarter of 2022, euro area headline inflation has been on a declining trajectory, with core inflation following suit in the second half of 2023. A key question is: How effectively did the QRF capture this disinflation path and provide real-time insights into the risks surrounding the Eurosystem's official projections?

Fig. 10 presents the observed year-on-year growth rates of headline HICP (left panel, blue line) and core inflation (right panel, blue line) from Q4 2022 to Q4 2024, the most recent inflation data available at the time of writing. These are shown alongside the Eurosystem projections for the respective quarters, produced at different horizons (ranging from two quarters ahead to the current quarter, indicated by different markers) and the 16th to 84th quantile range of the QRF forecasts at the time of the Eurosystem projections and in the intervening weeks (shaded orange areas).[23]

Fig. 10 illustrates the disinflation trajectory for headline and core inflation. Over the nine quarters covered in the picture, the QRF has demonstrated a high degree of accuracy. Notably, the width and shape of the QRF implied distribution have varied considerably over time, for example, around turning points, where the model has indicated greater uncertainty. For headline inflation, in the first two quarters of 2023, the disinflation process had only just begun to materialize, and both the QRF and the Eurosystem projections made two quarters ahead were still significantly higher than the final realized values. However, both forecasts adjusted relatively quickly over time. Since then, the QRF and Eurosystem projections have closely tracked the final outcomes as early as two quarters before the first release. Interestingly, the QRF has consistently signaled downside risks to the Eurosystem projections over the last five quarters in the sample, and these risks have generally materialized. A similar pattern emerged for core inflation. However, in the early stages of the disinflation process, both the QRF and the Eurosystem forecasts tended to underpredict core inflation, particularly at the two-quarter-ahead horizon. In the second half of 2023 and in the first quarter of 2024, much like for headline inflation, the QRF systematically signaled downside risks to the Eurosystem projections, which subsequently materialized. More recently, the QRF and Eurosystem core inflation projections have been closely aligned.

6. Conclusion

In this paper, we show that the quantile regression forest (QRF), a non-linear forecasting model, represents a valuable addition to the Eurosystem's inflation forecasting toolbox. Specifically, the QRF can enhance the assessment of risks surrounding the central tendency of the Eurosystem projections, which predominantly rely on linear models.

[22] See, in particular, chart 9 in that speech.

[23] The QRF forecasts are updated twice per month.

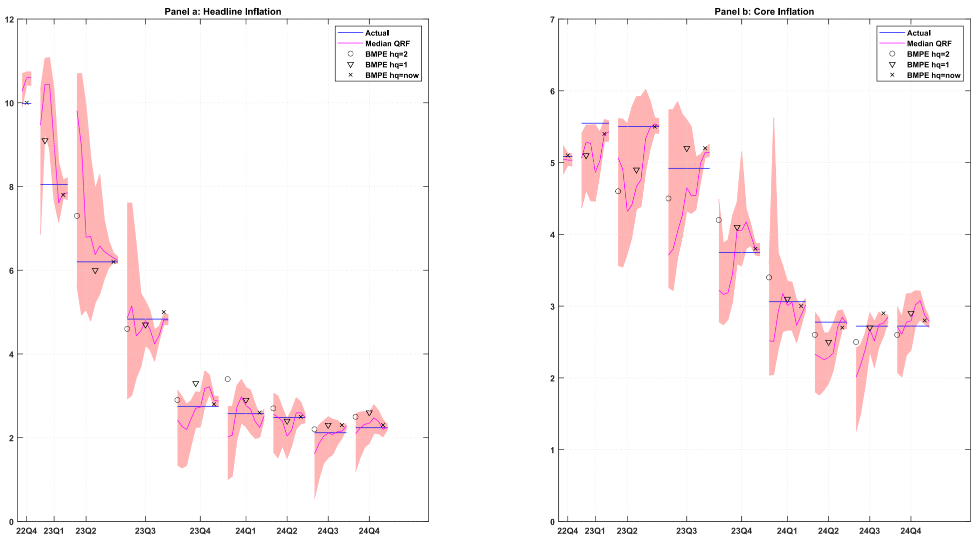

Fig. 10. Headline and core inflation, real-time forecasts 2022Q4–2024Q4.
Note: Blue solid lines: Headline (left panel) and Core (right panel) inflation outcomes, 2022Q4–2024Q4; Orange areas: 16th–84th range of QRF density forecasts, generally computed twice per month; Circle: Eurosystem projections, 2 quarters ahead; Triangle: Eurosystem projections, 1 quarter ahead; Star: Eurosystem projections, current quarter. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Our findings show that the QRF exhibits forecasting accuracy comparable to state-of-the-art linear and non-linear models for density forecasting. Furthermore, the QRF performs competitively against both institutional (BMPE) and survey-based (SPF) forecasts, despite the latter incorporating expert judgment.

The similar overall accuracy of linear and non-linear models over the full sample period suggests that the QRF serves as a complement rather than a substitute for the Eurosystem's existing modeling toolbox in inflation forecasting.

Regarding the broader question of whether euro area inflation exhibits non-linear dynamics, our analysis finds that non-linearity is more pronounced in core inflation than in headline inflation.

A final consideration pertains to the nature of the inflation projections in the Eurosystem (and in most central banks). The latter are effectively conditional forecasts, being based on a set of "assumptions" about the future path of relevant variables, such as commodity prices, interest rates and exchange rates. The QRF developed in this paper cannot handle conditional forecasts because, among other things, it does not model the joint distribution of inflation and its predictors. For example, Adrian et al. (2021) model the joint distribution of economic and financial conditions. We leave for future research the extension of the QRF in this direction.

Appendix A. Database of the QRF, SPF and BMPE

The target variables in our exercise are the euro area Harmonized Index of Consumer Prices (HICP) and the Harmonized Index of Consumer Prices excluding Energy and Food (HICPX). The former is what we consider our "headline" measure, while the latter is our "core" measure. Both measures and their dynamics are described at length in the main body of the paper. Table 5 reports the name (column 4) of the variables we use in our QRF and VARCOMB as predictors, the group of variables to which they belong (column 2) for the computation of the grouped Shapley values and the transformation we apply to make the variable stationary (column 3). Notice that our predictor set also includes two lags of inflation (which are included in the "past inflation" group in the analysis of the Shapley values).

A.1. Survey of professional forecasters

The Survey of Professional Forecasters (SPF) for the European Union has been taking place quarterly since the beginning of 1999. The survey asks to a panel of professional forecasters within the EU to give an estimate on the future values for euro area gross domestic product growth, HICP inflation, and the unemployment rate (de Vincent-Humphreys et al., 2019; Kenny et al., 2013). We focus here on inflation forecasts, for two separate horizons, namely current year and one-year-ahead.[24] The target for the two assessments are different. While the one-year-ahead concept (which, in the main body of the paper is defined as year-on-year inflation at h = 12 months ahead) measures the change in prices from the quarter preceding the assessment to four quarters later, the current year assessment pertains, in each quarter in which the survey is released, to the "average" inflation over the current year

[24] For each round, the target quarter refers to the current or one year after the latest official release available at the time of the questionnaire.

Table 5
Database - predictors.

	Group	Transf.	Variable description
1	G1	0	EU Commission, DG-ECFIN, Retail trade survey - expected business situation for 3 months ahead - Percentage balances
2	G1	0	EU Commission, DG-ECFIN, Consumer survey - financial situation over next 12 months - Percentages
3	G1	0	EU Commission, DG-ECFIN, Business climate indicator - Points of standard deviation
4	G1	0	EU Commission, DG-ECFIN, Consumer survey - general economic situation over next 12 months - Percentages
5	G1	0	EU Commission, DG-ECFIN, Consumer survey - major purchases over next 12 months - Percentages
6	G1	0	EU Commission, DG-ECFIN, Consumer survey - savings over next 12 months - Percentages
7	G1	0	EU Commission, DG-ECFIN, Consumer survey - unemployment expectations over next 12 months - Percentages
8	G1	0	EU Commission, DG-ECFIN, Consumer survey - consumer confidence indicator - Percentages
9	G1	0	EU Commission, DG-ECFIN, Economic sentiment indicator - Percentage balances
10	G1	0	EU Commission, DG-ECFIN, Industry survey - employment expectations for 3 months ahead - Percentage balances
11	G1	0	EU Commission, DG-ECFIN, Industry survey - production expectations for the 3 months ahead - Percentage balances
12	G6	0	EU Commission, DG-ECFIN, Industry survey - selling price expectations for the 3 months ahead - Percentage balances
13	G6	0	EU Commission, DG-ECFIN, Industry survey - selling price expectations for the months ahead, Intemediate Goods - Percentage balances
14	G6	0	EU Commission, DG-ECFIN, Industry survey - selling price expectations for the months ahead, Consumer Goods - Percentage balances
15	G6	0	EU Commission, DG-ECFIN, Consumer survey - price trends over next 12 months - Percentages
16	G7	0	ZEW, Short-term interest rate future (6 months) - Percentage balances
17	G1	0	Germany, ZEW, Economic situation future (6 months) - Percentage balances
18	G6	0	Germany, ZEW, Inflation rate future (6 months) - Percentage balances
19	G7	0	Germany, ZEW, Long-term interest rate future (6 months) - Percentage balances
20	G2	1	Equity/index - Baltic DRY Index (BDI) - Historical close, average of observations through period
21	G2	2	Bloomberg European Dated Brent Forties Oseberg Ekofisk (BFOE) Crude Oil Spot Price - Historical close - US dollar
22	G2	2	World market prices of raw materials, Index total, euro
23	G2	2	World market prices of raw materials, Index Total excluding energy, euro
24	G2	2	World market prices of raw materials, Energy, euro
25	G2	2	World market prices of raw materials, Crude oil, euro
26	G2	2	World market prices of raw materials, Industrial raw materials, euro
27	G2	2	World market prices of raw materials, Food and tropical beverages, euro
28	G2	2	ECB Commodity Price index Euro denominated, import weighted, Non-food
29	G2	2	ECB Commodity Price index Euro denominated, import weighted, Agricultural raw materials
30	G2	0	EXCH.RATE: US DOLLARS/1 EUR,SPOT AT 2:15 PM (CET) D,W,M,Q,A-AVG
31	G2	0	ECB Nominal effective exch. rate of the Euro against, EER-12 group of trading partners: AU,CA,DK,HK,JP,NO,SG,KR,SE,CH,GB,US,EA excluding the Euro
32	G3	2	Producer Price Index, domestic sales, Consumer goods industry
33	G3	2	Producer Price Index, domestic sales, MIG Durable Consumer Goods Industry
34	G3	2	Producer Price Index, domestic sales, MIG Non-durable Consumer Goods Industry
35	G3	2	Producer Price Index, domestic sales, MIG Intermediate Goods Industry
36	G3	2	Producer Price Index, domestic sales, MIG Capital Goods Industry
37	G3	2	Producer Price Index, domestic sales, MIG Energy
38	G3	2	Producer Price Index, domestic sales, MANUFACTURING
39	G3	2	Producer Price Index, domestic sales, Total Industry (excluding construction)
40	G4	0	Indicator of negotiated wage rates, Total - Annual rate of change
41	G4	0	Indicator of negotiated wage rates - total excluding bonuses, Total - Annual rate of change
42	G1	2	Industrial Production Index, Total Industry (excluding construction)
43	G1	1	Building Permits/dwellings, Residential buildings except residences for communities
44	G1	0	European Labour Force Survey; Unemployment rate; Total; Age 15 to 74
45	G1	1	EA19 Leading Indicators OECD > Leading indicators > CLI > Amplitude adjusted/Level. rate or national currency
46	G1	0	United States; European Labour Force Survey; Unemployment rate; Total; Age 15 to 74
47	G5	0	Euribor 3-month - Last trade price or value
48	G5	0	Benchmark bond - Euro area 10-year Government Benchmark bond yield - Yield
49	G5	2	European Monetary Union Market Index. Equity Index.
50	G5	0	IBES MSCI EMU Index Earnings. Weighted average long term growth EPS (Earnings per share) forecast expressed as a percentage

(*continued on next page*)

(average means, roughly, the average of the four year-on-year inflation rates in the four quarter of the current year). Respondents are asked to give both a point forecast and to assign probabilities for each variable's future outcome falling within pre-determined ranges. The individual responses are then aggregated, and a histogram of average probabilities for the economic outlook results. We do not focus on individual responses, following the results in Genre et al. (2013), where the simple average is proven to be

Table 5 (*continued*).

51	G5	2	Euro area - Equity/index - European Monetary Union Consumer Goods Index (EUR)
52	G5	2	Equity/index - European Monetary Union Consumer Services Index (EUR) - Historical close
53	G5	2	Monetary aggregate M1
54	G5	2	Monetary aggregate M3
55	G5	2	Monetary aggregate M2
56	G5	2	Loans, Total maturity, All currencies combined - Euro area (changing composition) counterpart
57	G6	2	US - CONSUMER PRICES, ALL ITEMS (ALL URBAN CONSUMERS)
58	G6	2	US - CONSUMER PRICES, CORE INFLATION (URBAN CONSUMERS)
59	G6	0	EU Commission, DG-ECFIN, Services survey - selling price expectations for the 3 months ahead - Percentage balances

Note: G1: real activity, G2 : commodity prices, G3: PPI, G4: wages, G5: financial, G6: inflation expectations, G7: interest rate expectations. Transformations for stationarity: 0 = no transformation, 1 = natural logarithm, 2 = first difference of natural logarithm.

the best combination method. Other aggregation methods include optimal pooling like in Conflitti et al. (2015), and a more recent work by Diebold et al. (2020), where the authors propose to build regularized mixtures of individual densities.

The first SPF vintage for headline inflation corresponds to February 2002.[25] The forecasts for those vintages are supposed to be produced around mid- of the previous month. We assume that forecasters had data up to the previous December when matching the information provided to the QRF. Then, we continue to update the estimation sample by adding one quarter (effectively, three months) at a time, and we repeat all the steps of the forecasting exercise until exhaustion of the sample.

A.2. Inflation projections from the BMPE

Eurosystem and ECB staff produce macroeconomic projections (BMPE) that cover the outlook for the euro area and the wider global economy. These contribute to the ECB Governing Council's assessment of economic developments and risks to price stability.

They are published four times a year (in March, June, September and December).

The first BMPE vintage corresponds to March 2002. We assume that forecasters had data up to January 2002 when matching the information provided to the QRF. Then, we continue to update the estimation sample by adding one quarter (effectively, three months) at a time, and we repeat all the steps of the forecasting exercise until exhaustion of the sample.

A.3. Fitting the skew-t distribution

While we generally use only the empirical prediction quantiles for our out-of-sample evaluations, to evaluate the confidence regions of the test of Rossi and Sekhposyan (2019) and to compute exact quantiles for the SPF to match those of the other models, we need to fit a distribution.

The skew-t distribution is a flexible, parametric density that allows for fat tails as well as asymmetries, controlled by the parameters defining the distribution.

We define the skew-t (ST) for a variable Y as:

$$Y \sim ST(\xi, \omega, \alpha, \nu)$$

where ξ is a location parameter, ω is the scale, α is the slant parameter that determines the skewness of the distribution, and ν is the degrees of freedom.

In order to fit a skew-t to our density forecasts, we match the empirical quantiles of our forecasts. The only set of forecasts for which it may be challenging to derive the empirical quantiles, is the SPF. We derive the empirical quantiles from the SPF histograms. Specifically, we consider, for each release of the SPF, the histogram based on the reported probabilities, for the horizons of interest, i.e. for the forecasts of the current year HICP and HICPX inflation and of the year-on-year growth rates of HICP and HICPX one year ahead. We obtain the quantiles of the empirical cdf from the bin edges of the SPF histogram, and we match the closest possible quantiles to the quantiles we used for the QRF.

Once we have the quantiles, we follow Montes Galdon et al. (2023) and fit the pdf of a skew-t distribution.[26] Note however that we need to keep the degrees of freedom of the distribution, ν, as a discrete value. Therefore, we proceed as follows. We construct first a grid for the degrees of freedom. For each value of the grid, we find the location, scale and slant parameters with the best match of the quantiles provided from the pdf. At this stage, we have a set of parameters matching the quantiles we have chosen, given a certain value of ν. In this set, we select the parameters with the minimum squared 2-norm distance from the empirical quantiles.

Appendix B. Comparison of RMSE between median QRF and random walk forecasts

This comparison is conducted at the monthly frequency, which is available for the three models involved (see Table 6).

[25] The first vintage available for core inflation is February 2017. Given the short sample available, we exclude SPF core inflation predictions from the analysis.

[26] Notice that there are alternative approaches as in Engelberg et al. (2009), which assumes a normal or a beta distribution for the SPF histograms and Billio et al. (2013), which produces a continuous SPF distribution, as well as draws from this distribution, using a kernel smoother.

Table 6
RMSE of QRF and RW for headline and core inflation.

Horizon	RW	QRF mean
Panel a: Headline inflation		
h = 3	1.23	0.97
h = 6	1.14	0.99
h = 9	1.04	0.99
h = 12	0.97	0.99
Panel b: Core inflation		
h = 3	1.41	1.09
h = 6	1.21	1.04
h = 9	0.99	1.00
h = 12	0.91	1.01

Note: Column 2: RW; Column 3: QRF mean. A value larger than one indicates that the QRF median forecasts are more accurate than a specific model (indicated in the column) at a specific horizon (indicated in the row).

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.euroecorev.2025.105079.

Data availability

The data we used are already included in our replication files. Moreover the results of our recursive out-of-sample analysis with the four econometric models are also available.

References

Aas, K., Jullum, M., Løland, A., 2020. Explaining individual predictions when features are dependent: More accurate approximations to Shapley values. arXiv:1903.10464.

Adams, P.A., Adrian, T., Boyarchenko, N., Giannone, D., 2021. Forecasting macroeconomic risks. Int. J. Forecast. 37 (3), 1173–1191, URL https://ideas.repec.org/a/eee/intfor/v37y2021i3p1173-1191.html.

Adrian, T., Boyarchenko, N., Giannone, D., 2019. Vulnerable growth. Am. Econ. Rev. 109 (4), 1263–1289, URL https://ideas.repec.org/a/aea/aecrev/v109y2019i4p1263-89.html.

Adrian, T., Boyarchenko, N., Giannone, D., 2021. Multimodality in macrofinancial dynamics. Internat. Econom. Rev. 62 (2), 861–886, URL https://ideas.repec.org/a/wly/iecrev/v62y2021i2p861-886.html.

Akerlof, G.A., Dickens, W.R., Perry, G.L., 1996. The macroeconomics of low inflation. Brook. Pap. Econ. Act. 27 (1), 1–76, URL https://ideas.repec.org/a/bin/bpeajo/v27y1996i1996-1p1-76.html.

Amburgey, A., McCracken, M.W., 2023a. Growth-at-Risk is Investment-at-Risk. Working Papers 2023–020, Federal Reserve Bank of St. Louis, URL https://ideas.repec.org/p/fip/fedlwp/96594.html.

Amburgey, A.J., McCracken, M.W., 2023b. On the real-time predictive content of financial condition indices for growth. J. Appl. Econometrics 38 (2), 137–163, URL https://ideas.repec.org/a/wly/japmet/v38y2023i2p137-163.html.

Athey, S., Tibshirani, J., Wager, S., 2016. Generalized Random Forests. Papers 1610.01271, arXiv.org, URL https://ideas.repec.org/p/arx/papers/1610.01271.html.

Atkeson, A., Ohanian, L.E., 2001. Are phillips curves useful for forecasting inflation? Q. Rev. 25 (Win), 2–11, URL https://ideas.repec.org/a/fip/fedmqr/y2001iwinp2-11nv.25no.1.html.

Avramov, D., 2002. Stock return predictability and model uncertainty. J. Financ. Econ. 64, 423–458.

Azzalini, A., Capitanio, A., 2003. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. J. R. Stat. Soc. Ser. B 65, 367–389.

Bai, J., Ng, S., 2009. Boosting diffusion indices. J. Appl. Econometrics 24 (4), 607–629, URL https://ideas.repec.org/a/jae/japmet/v24y2009i4p607-629.html.

Banbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. J. Appl. Econometrics 25 (1), 71–92, URL https://ideas.repec.org/a/jae/japmet/v25y2010i1p71-92.html.

Banbura, M., Lenza, M., Paredes, J., 2024. Forecasting inflation in the US and in the euro area. In: Michael P. Clements, Ana Beatriz Galvão (Eds.), Handbook of Research Methods and Applications in Macroeconomic Forecasting. Chapter 9, pp. 218–245.

Benigno, P., Eggertsson, G.B., 2023. It's Baaack: The Surge in Inflation in the 2020s and the Return of the Non-Linear Phillips Curve. NBER Working Papers 31197, National Bureau of Economic Research, Inc, URL https://ideas.repec.org/p/nbr/nberwo/31197.html.

Benigno, P., Ricci, L.A., 2011. The inflation-output trade-off with downward wage rigidities. Am. Econ. Rev. 101 (4), 1436–1466, URL https://ideas.repec.org/a/aea/aecrev/v101y2011i4p1436-66.html.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. J. Bus. Econom. Statist. 19 (4), 465–474, URL https://ideas.repec.org/a/bes/jnlbes/v19y2001i4p465-74.html.

Billio, M., Casarin, R., Ravazzolo, F., van Dijk, H.K., 2013. Time-varying combinations of predictive densities using nonlinear filtering. J. Econometrics 177, 213–232.

Bobeica, E., Jarociński, M., 2019. Missing disinflation and missing inflation: A VAR perspective. Int. J. Central Bank. 15 (1), 199–232, URL https://ideas.repec.org/a/ijc/ijcjou/y2019q1a5.html.

Botelho, V., Foroni, C., Renzetti, A., 2023. Labour at Risk. Working Paper Series 2840, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20232840.html.

Boyarchenko, N., Crump, R.K., Elias, L., Gaffney, I.L., 2023. Look Out for Outlook-at-Risk. Liberty Street Economics 20230517, Federal Reserve Bank of New York, URL https://ideas.repec.org/p/fip/fednls/96162.html.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Buckmann, M., Joseph, A., 2022. An Interpretable Machine Learning Workflow with an Application to Economic Forecasting. Bank of England Working Papers 984, Bank of England.

Cavallo, A., Lippi, F., Miyahara, K., 2023. Large Shocks Travel Fast. NBER Working Papers 31659, National Bureau of Economic Research, Inc, URL https://ideas.repec.org/p/nbr/nberwo/31659.html.

Chavleishvili, S., Kremer, M., 2023. Measuring Systemic Financial Stress and its Risks for Growth. Working Paper Series 2842, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20232842.html.

Chavleishvili, S., Kremer, M., Lund-Thomsen, F., 2023. Quantifying Financial Stability Trade-Offs for Monetary Policy: A Quantile VAR Approach. Working Paper Series 2833, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20232833.html.

Chavleishvili, S., Manganelli, S., 2019. Forecasting and Stress Testing with Quantile Vector Autoregression. Working Paper Series 2330, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20192330.html.

Clark, T.E., Huber, F., Koop, G., Marcellino, M., 2022. Forecasting US Inflation Using Bayesian Nonparametric Models. Working Papers 22–05, Federal Reserve Bank of Cleveland, URL https://ideas.repec.org/p/fip/fedcwq/93787.html.

Clark, T.E., Huber, F., Koop, G., Marcellino, M., Pfarrhofer, M., 2021. Tail Forecasting with Multivariate Bayesian Additive Regression Trees. Working Papers 21-08R, Federal Reserve Bank of Cleveland, URL https://ideas.repec.org/p/fip/fedcwq/90366.html.

Conflitti, C., De Mol, C., Giannone, D., 2015. Optimal combination of survey forecasts. Int. J. Forecast. 31 (4), 1096–1103.

Corradi, V., Swanson, N.R., 2006. Predictive density evaluation. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting. In: Handbook of Economic Forecasting, vol. 1, Elsevier, pp. 197–284, URL https://ideas.repec.org/h/eee/ecofch/1-05.html. chapter 5.

Costain, J., Nakov, A., Petit, B., 2022. Flattening of the Phillips curve with state-dependent prices and wages. Econ. J. 132, 546–581.

Crump, R.K., Everaert, M., Giannone, D., Hundtofte, S., 2018. Changing Risk-Return Profiles. Staff Reports 850, Federal Reserve Bank of New York, URL https://ideas.repec.org/p/fip/fednsr/850.html.

Darracq Pariès, M., Notarpietro, A., Kilponen, J., Papadopoulou, N., Zimic, S., Aldama, P., Langenus, G., Alvarez, L.J., Lemoine, M., Angelini, E., 2021. Review of Macroeconomic Modelling in the Eurosystem: Current Practices and Scope for Improvement. Occasional Paper Series 267, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbops/2021267.html.

de Bondt, G., Hahn, E., Zekaite, Z., 2018. ALICE: A New Inflation Monitoring Tool. Working Paper Series 2175, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20182175.html.

de Vincent-Humphreys, R., Dimitrova, I., Falck, E., Henkel, L., 2019. Twenty years of the ECB survey of professional forecasters. Econ. Bull. Artic. 1.

Del Negro, M., Lenza, M., Primiceri, G., Tambalotti, A., 2020. What's up with the Phillips curve? Brook. Pap. Econ. Act. Spring.

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. Internat. Econom. Rev. 39 (4), 863–883, URL https://ideas.repec.org/a/ier/iecrev/v39y1998i4p863-83.html.

Diebold, F.X., Shin, M., Zhang, B., 2020. On the aggregation of probability assessments: Regularized mixtures of predictive densities for eurozone inflation and real interest rates. arXiv preprint arXiv:2012.11649.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. Econometric Rev. 3 (1), 1–100.

Ehrmann, M., Ferrucci, G., Lenza, M., O'Brien, D., 2018. Measures of underlying inflation for the euro area. Econ. Bull. Artic. 4, URL https://ideas.repec.org/a/ecb/ecbart/201800043.html.

Elliott, G., Timmermann, A., 2016. Economic Forecasting. Princeton University Press.

Engelberg, J., Manski, C.F., Williams, J., 2009. Comparing the point predictions and subjective probability distributions of professional forecasters. J. Bus. Econom. Statist. 27 (1), 30–41.

Fahr, S., Smets, F., 2010. Downward wage rigidities and optimal monetary policy in a monetary union. Scand. J. Econ. 112 (4), 812–840, URL https://ideas.repec.org/a/bla/scandj/v112y2010i4p812-840.html.

Faust, J., Gilchrist, S., Wright, J.H., Zakrajšsek, E., 2013. Credit spreads as predictors of real-time economic activity: A Bayesian model-averaging approach. Rev. Econ. Stat. 95 (5), 1501–1519, URL https://ideas.repec.org/a/tpr/restat/v95y2013i5p1501-1519.html.

Faust, J., Wright, J.H., 2013. Chapter 1 - forecasting inflation. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting. In: Handbook of Economic Forecasting, vol. 2, Elsevier, pp. 2–56, URL https://www.sciencedirect.com/science/article/pii/B9780444536839000013.

Fernandez, C., Ley, E., Steel, M.F.J., 2001. Model uncertainty in cross-country growth regressions. J. Appl. Econometrics 16 (5), 563–576, URL https://ideas.repec.org/a/jae/japmet/v16y2001i5p563-576.html.

Ferrara, L., Marcellino, M., Mogliani, M., 2015. Macroeconomic forecasting during the great recession: The return of non-linearity? Int. J. Forecast. 31 (3), 664–679, URL https://ideas.repec.org/a/eee/intfor/v31y2015i3p664-679.html.

Figueres, J.M., Jarociński, M., 2020. Vulnerable growth in the euro area: Measuring the financial conditions. Econom. Lett. 191 (C), URL https://ideas.repec.org/a/eee/ecolet/v191y2020ics016517652030104x.html.

Forbes, K.J., Gagnon, J.E., Collins, C.G., 2021. Low Inflation Bends the Phillips Curve Around the World: Extended Results. Working Paper Series WP21-15, Peterson Institute for International Economics, URL https://ideas.repec.org/p/iie/wpaper/wp21-15.html.

Genre, V., Kenny, G., Meyler, A., Timmermann, A., 2013. Combining expert forecasts: Can anything beat the simple average? Int. J. Forecast. 29 (1), 108–121.

Giacomini, R., Rossi, B., 2009. Detecting and predicting forecast breakdowns. Rev. Econ. Stud. 76 (2), 669–705, URL https://ideas.repec.org/a/oup/restud/v76y2009i2p669-705.html.

Giacomini, R., Rossi, B., 2010. Forecast comparisons in unstable environments. J. Appl. Econometrics 25 (4), 595–620, URL https://ideas.repec.org/a/jae/japmet/v25y2010i4p595-620.html.

Giannone, D., Lenza, M., Momferatou, D., Onorante, L., 2014. Short-term inflation projections: A Bayesian vector autoregressive approach. Int. J. Forecast. 30 (3), 635–644, URL https://ideas.repec.org/a/eee/intfor/v30y2014i3p635-644.html.

Giannone, D., Lenza, M., Primiceri, G.E., 2015. Prior selection for vector autoregressions. Rev. Econ. Stat. 97 (2), 436–451, URL https://ideas.repec.org/a/tpr/restat/v97y2015i2p436-451.html.

Giannone, D., Lenza, M., Primiceri, G.E., 2021. Economic predictions with big data: The illusion of sparsity. Econometrica 89 (5), 2409–2437, URL https://ideas.repec.org/a/wly/emetrp/v89y2021i5p2409-2437.html.

Giannone, D., Primiceri, G., 2024. The Drivers of Post-Pandemic Inflation. CEPR Discussion Papers 19377, C.E.P.R. Discussion Papers, URL https://ideas.repec.org/p/cpr/ceprdp/19377.html.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. J. R. Stat. Soc. Ser. B 69 (2), 243–268, URL https://EconPapers.repec.org/RePEc:bla:jorssb:v:69:y:2007:i:2:p:243-268.

Gneiting, T., Raftery, A.E., 2007. Strict proper scoring rules, prediction and estimation. J. Amer. Statist. Assoc. 102, 359–378.

González-Rivera, G., Sun, Y., 2015. Generalized autocontours: Evaluation of multivariate density models. Int. J. Forecast. 31 (3), 799–814, URL https://ideas.repec.org/a/eee/intfor/v31y2015i3p799-814.html.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., Surprenant, S., 2022. How is machine learning useful for macroeconomic forecasting? J. Appl. Econometrics 37 (5), 920–964, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/jae.2910. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2910.

Hamill, T., 2000. Interpretation of rank histograms for verifying ensemble forecasts. Mon. Weather Rev. 129.

Hong, Y., Li, H., Zhao, F., 2007. Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates. J. Econometrics 141 (2), 736–776, URL https://ideas.repec.org/a/eee/econom/v141y2007i2p736-776.html.

Inoue, A., 2001. Testing for distributional change in time series. Econometric Theory 17 (1), 156–187, URL https://ideas.repec.org/a/cup/etheor/v17y2001i01p156-187_17.html.

Inoue, A., Kilian, L., 2008. How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation. J. Amer. Statist. Assoc. 103, 511–522, URL https://ideas.repec.org/a/bes/jnlasa/v103y2008mjunep511-522.html.

Jin, S., Su, L., Ullah, A., 2014. Robustify financial time series forecasting with bagging. Econometric Rev. 33 (5–6), 575–605, URL https://ideas.repec.org/a/taf/emetrv/v33y2014i5-6p575-605.html.

Kenny, G., Kostka, T., Masera, F., 2013. Can Macroeconomists Forecast Risk? Event-Based Evidence from the Euro Area SPF. ECB Working Paper.

Kiley, M.T., 2022. Unemployment risk. J. Money Credit. Bank. 54 (5), 1407–1424, URL https://ideas.repec.org/a/wly/jmoncb/v54y2022i5p1407-1424.html.

Knüppel, M., 2015. Evaluating the calibration of multi-step-ahead density forecasts using raw moments. J. Bus. Econom. Statist. 33 (2), 270–281, URL https://ideas.repec.org/a/taf/jnlbes/v33y2015i2p270-281.html.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50, URL http://www.jstor.org/stable/1913643.

Koester, G., Lis, E., Nickel, C., Osbat, C., Smets, F., 2021. Understanding Low Inflation in the Euro Area from 2013 to 2019: Cyclical and Structural Drivers. Occasional Paper Series 280, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbops/2021280.html.

Korobilis, D., Landau, B., Musso, A., Phella, A., 2021. The Time-Varying Evolution of Inflation Risks. Working Paper Series 2600, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20212600.html.

Kuik, F., Adolfsen, J.F., Lis, E.M., Meyler, A., 2022. Energy price developments in and out of the COVID-19 pandemic – from commodity prices to consumer prices. ECB Econ. Bull. Issue 4/2022.

Lane, P., 2024. Disinflation in the euro area. Speech by Philip R. Lane, member of the Executive Board of the ECB, Hutchins Center on Fiscal Monetary Policy at the Brookings Institution Washington, D.C., 8 February 2024 available at https://www.ecb.europa.eu/press/key/date/2024/html/ecb.sp240208~e775b07928.en.html.

Lenza, M., 2011. Revisiting the information content of core inflation. Res. Bull. 14, 11–13, URL https://ideas.repec.org/a/ecb/ecbrbu/201100143.html.

Lenza, M., Moutachaker, I., Paredes, J., 2023. Density Forecasts of Inflation: A Quantile Regression Forest Approach. Working Paper Series 2830, European Central Bank, URL https://ideas.repec.org/p/ecb/ecbwps/20232830.html.

Lindé, J., Trabandt, M., 2019. Resolving the Missing Deflation Puzzle. CEPR Discussion Papers 13690, C.E.P.R. Discussion Papers, URL https://ideas.repec.org/p/cpr/ceprdp/13690.html.

Litterman, R.B., 1979. Techniques of Forecasting Using Vector Autoregressions. Tech. Rep..

López-Salido, J.D., Loria, F., 2020. Inflation at Risk. Finance and Economics Discussion Series 2020–013, Board of Governors of the Federal Reserve System (U.S.), URL https://ideas.repec.org/p/fip/fedgfe/2020-13.html.

Lundberg, S.M., Erion, G.G., Lee, S.-I., 2019. Consistent individualized feature attribution for tree ensembles. arXiv:1802.03888.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems. Vol. 30, Curran Associates, Inc., URL https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Medeiros, M.C., Vasconcelos, G.F.R., Veiga, Á., Zilberman, E., 2021. Forecasting inflation in a data-rich environment: The benefits of machine learning methods. J. Bus. Econom. Statist. 39 (1), 98–119, arXiv:https://doi.org/10.1080/07350015.2019.1637745.

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Montes Galdon, C., Paredes, J., Wolf, E., 2022. Conditional Density Forecasting: A Tempered Importance Sampling Approach. ECB Working Paper 2754.

Montes Galdon, C., Paredes, J., Wolf, E., 2023. A Robust Approach to Tilting: Parametric Relative Entropy. Tech. Rep., Mimeo.

Ng, S., 2013. Variable selection in predictive regressions. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Economic Forecasting. In: Handbook of Economic Forecasting, vol. 2, Elsevier, pp. 752–789, URL https://ideas.repec.org/h/eee/ecofch/2-752.html. chapter 0.

Probst, P., Wright, M.N., Boulesteix, A.-L., 2019. Hyperparameters and tuning strategies for random forest. WIREs Data Min. Knowl. Discov. 9 (3), e1301, arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1301. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1301.

Rapach, D., Strauss, J., 2010. Bagging or combining (or both)? An analysis based on forecasting U.S. employment growth. Econometric Rev. 29 (5–6), 511–533, URL https://ideas.repec.org/a/taf/emetrv/v29y2010i5-6p511-533.html.

Rossi, B., Sekhposyan, T., 2016. Forecast rationality tests in the presence of instabilities, with applications to federal reserve and survey forecasts. J. Appl. Econometrics 31 (3), 507–532, URL https://ideas.repec.org/a/wly/japmet/v31y2016i3p507-532.html.

Rossi, B., Sekhposyan, T., 2019. Alternative tests for correct specification of conditional predictive densities. J. Econometrics 208 (2), 638–657, URL https://ideas.repec.org/a/eee/econom/v208y2019i2p638-657.html.

Sala-I-Martin, X., Doppelhofer, G., Miller, R.I., 2004. Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. Am. Econ. Rev. 94 (4), 813–835, URL https://ideas.repec.org/a/aea/aecrev/v94y2004i4p813-835.html.

Shapley, L.S., 1952. A Value for N-Person Games. RAND Corporation, Santa Monica, CA.

Strumbelj, E., Kononenko, I., 2010. An efficient explanation of individual classifications using game theory. J. Mach. Learn. Res. 11, 1–18.

Varian, H.R., 2014. Big data: New tricks for econometrics. J. Econ. Perspect. 28 (2), 3–28, URL https://ideas.repec.org/a/aea/jecper/v28y2014i2p3-28.html.

Wager, S., Athey, S., 2018. Estimation and inference of heterogeneous treatment effects using random forests. J. Amer. Statist. Assoc. 113 (523), 1228–1242, URL https://ideas.repec.org/a/taf/jnlasa/v113y2018i523p1228-1242.html.

Wright, J.H., 2009. Forecasting US inflation by Bayesian model averaging. J. Forecast. 28 (2), 131–144, URL https://ideas.repec.org/a/jof/jforec/v28y2009i2p131-144.html.