EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Michele Lenza, Inès Moutachaker, Joan Paredes

Density forecasts of inflation: a quantile regression forest approach

No 2830

Abstract

Density forecasts of euro area inflation are a fundamental input for a medium-term oriented central bank, such as the European Central Bank (ECB). We show that a quantile regression forest, capturing a general non-linear relationship between euro area (headline and core) inflation and a large set of determinants, is competitive with state-of-the-art linear benchmarks and judgemental survey forecasts. The median forecasts of the quantile regression forest are very collinear with the ECB point inflation forecasts, displaying similar deviations from "linearity". Given that the ECB modelling toolbox is overwhelmingly linear, this finding suggests that the expert judgement embedded in the ECB forecast may be characterized by some mild non-linearity.

Keywords: Inflation, Non-linearity, Quantile Regression Forest.

JEL Codes: C52, C53, E31, E37

Non-Technical Summary

Density inflation forecasts are essential inputs for monetary policy decisions. In this paper, we aim to enrich the essentially linear Eurosystem forecasting toolbox by developing a new forecasting model for headline and core inflation (proxied by the growth rates of HICP and HICP excluding energy and food prices) which is able to capture non-linear inflation dynamics and to deliver density forecasts.

The model we propose is borrowed from the machine learning literature and is defined as the Quantile Regression Forest (QRF), developed in Meinshausen (2006), which is a variant of the popular Random Forest of Breiman (2001). As potential determinants of inflation, in our model, we choose a set of 60 variables inspired by the Phillips Curve framework, encompassing measures of inflation expectations, cost pressures, real activity and financial variables.

We compare our QRF density forecasts to those from a state-of-the-art linear model (a combination of BVAR models, defined as VARCOMB) and the ECB Survey of Professional Forecasters (SPF). We find that the QRF density forecasts are competitive with those from the benchmarks. Over our evaluation sample, the QRF forecasts are outperformed by those from VARCOMB during and in the aftermath of the Great Recession, while they are better equipped to capture the prolonged period of low inflation before and during the COVID pandemic. In general, the relative accuracy of the QRF with respect to the benchmarks is higher for core than for headline inflation. We also study which predictors are "more relevant" for inflation in the QRF and which type of functional forms are better suited to capture the relationship between inflation and its predictors. We find a non-linear predictive relationship with measures of inflation expectations, particularly for core inflation.

Overall, we conclude that the QRF is a valid addition to the Eurosystem forecasting toolbox, although it is better seen as a complement rather than a substitute for existing techniques. Inflation dynamics seem to be characterized by a mild non-linearity, which is more evident for inflation in the HICP excluding food and energy, our measure of core inflation. These results imply that the dynamics in the energy and food components are broadly characterized by linear dynamics and may over-shadow the mild non-linearity in core in period of high volatility of commodity prices.

We also compare our median QRF forecasts to the Eurosystem point inflation forecasts, and we find that the two sets of forecasts display similar "deviations from linearity", roughly measured as the gaps of the two sets of forecasts with respect to the median VARCOMB forecasts. Given the overwhelming linearity of the Eurosystem forecasting toolbox, this implies that the judgmental component of the Eurosystem projections tends to embed a mild non-linearity in the inflation projections.

1 Introduction

The mandate of the European Central Bank (ECB) is to maintain price stability *over the medium term*. The medium term orientation implies that sources of fluctuations with temporary effects on inflation are more likely to be looked through, while driving forces of a more persistent nature may have an impact on monetary policy decisions. For this reason, the inflation projections, which condense the views of the Eurosystem staff on inflation dynamics, are a crucial input for monetary policy decisions. At the same time, the economic projections are surrounded by uncertainty and policy decisions hinge on a careful characterization of the likelihood of different potential current and future scenarios, carried out in the so called "risk assessment", rather than merely on point forecasts.

Modelling inflation dynamics in the euro area remains elusive, owing to the many potential driving factors of inflation and the difficulty in capturing their relationship with inflation dynamics (see, for example, Koester et al., 2021). One key point is whether inflation dynamics are better characterized by a linear or a non-linear relationship with its potentially many determinants. The thorough description of the Eurosystem economic analysis in Darracq Pariès et al. (2021), carried out for the recent ECB strategy review, reveals that the Eurosystem modeling toolbox consists mostly of linear models. Yet, the relationship of consumer prices with their determinants may be characterized by non-linearity. For example, among others, Lindé and Trabandt (2019) and Costain et al. (2022) argue that some form of state-dependence could help to explain the inflation dynamics over the last fifteen years. Moreover, recent literature has highlighted how the density of economic variables such as GDP and inflation may be characterized by relevant non-linearities. For example, Adrian et al. (2019) shows that the upper and lower quantiles of the GDP predictive density are affected by different variables and Carriero et al. (2016) and López-Salido and Loria (2020) argue that similar considerations also concern inflation predictions.

In this paper, we introduce a novel non-linear model for the density forecast of euro area inflation. We measure headline inflation as the rate of change of the Harmonized Index of Consumer Prices (HICP), and we also look at a measure of "core inflation", i.e. the rate of change of the Harmonized Index of Consumer Prices excluding Energy and Food prices (HICPex).[1] The potential determinants of headline and core inflation in our model are broadly inspired by the Phillips Curve framework and include measures of inflation expectations, cost pressures, real activity and financial variables. To model the relationship of the determinants with inflation, we employ the quantile regression forest (QRF) of Meinshausen (2006), a variant of the random forest of Breiman (2001), which allows us to obtain a density forecast. The quantile regression forest and the random forest are appealing because they are able to capture very general functional forms, encompassing non-linearity in inflation dynamics.

The random forest is an ensemble technique, combining a number of non-linear predictive models, called regression trees. Regression trees split the sample of the predictors in (potentially many)

[1] In the rest of the paper, we refer to HICPex inflation as to core inflation. However, despite the popularity of "exclusion measures" of consumer prices to proxy core inflation, there are many other measures of core inflation, with advantages and shortcomings, and policy-makers in practice look at a range of different measures rather than settling on a specific one. For more details, see for example Lenza (2011) and Ehrmann et al. (2018).

sub-samples (called "leaves") and predict the target variable by means of the average value of the latter in each particular sub-sample. Quantile regression forests are based on the same principles but entail the estimation of empirical quantiles of the distribution of the target variables in the leaves and, hence, are able to handle density forecasting. These models can capture very general forms of non-linearity because they do not assume any specific parametric relationship between predictors and the target variables.

In our empirical application, we evaluate the QRF in a recursive out-of-sample exercise spanning the last twenty years, ending in December 2022. We focus on density forecasts up to the one-year ahead horizon and we compare the accuracy of the QRF predictions with those from a state-of-the-art linear benchmark (a combination of a large number of Bayesian VAR models, VARCOMB) and survey forecasts (the ECB Survey of Professional Forecasters, SPF in short). We also carry out a comparison of the median QRF prediction with institutional forecasts (the published Eurosystem Inflation Projections, BMPE in short); for this comparison, we use only the median QRF because the institutional forecasts are publicly available in terms of point forecasts. We look at the tests of correct calibration of the forecasts developed in Rossi and Sekhposyan (2019) and, then, we compare the forecasts by using a proper scoring rule, i.e. the continuous ranked probability score (CRPS) of Gneiting and Raftery (2007), which assess both calibration and sharpness and allows us to rank the different forecasting methods.

We find that the QRF produces well calibrated density forecasts for euro area inflation and it is competitive with state-of-the-art linear and survey forecasting methods, for headline and, in particular, core inflation. The predictive performance of the QRF is particularly good at horizons of up to six months, while it deteriorates in comparison with linear models and survey forecasts at the longer horizons. For example, for the assessment of current year inflation conducted over the last two quarters of the year, the QRF density forecasts are much sharper around the true value than those from the SPF. We also find that the relative forecasting performance of the QRF and the VARCOMB changes over time and, while the QRF did not perform too well during and in the aftermath of the Great Financial Crisis of 2007-2009, it outperforms VARCOMB over the protracted period of low inflation which characterized the euro area before the COVID pandemic. These results suggest that euro area inflation dynamics may be characterized by non-linearity to a certain extent, although the differences in accuracy with respect to VARCOMB are relatively small and, hence, the non-linearity is relatively mild.

Interestingly, the evidence of non-linearity seems to be stronger for core than headline inflation. Given that our measure of core inflation is a sub-component of headline inflation excluding the energy and the food components, the evidence of non-linearity for the energy and food prices is much less compelling. The stronger non-linearity of core inflation is further qualified when we study the contribution of different predictors to the inflation forecasts, by means of the Shapley values (see Shapley, 1952; Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017; Buckmann and Joseph, 2022). Indeed, the analysis of the functional forms relating the different predictors to inflation reveals that most of the relevant predictors are in an essentially linear relationship with inflation, except for measures of inflation expectations which display a state-dependent relationship, especially with core inflation. We conclude that the QRF is a useful addition to the Eurosystem forecasting toolbox but, at the same time, it should be considered more as a complement, rather than a substitute of the existing methods to forecast inflation.

Turning to the comparison of the QRF median forecasts with the judgemental Eurosystem inflation forecasts, the two sets of forecasts have similar accuracy and display very similar dynamics. In particular, when we look at the gaps between our median VARCOMB forecasts with the median QRF and the Eurosystem forecasts, a rough measure of "distance from linearity", the gaps are quite strongly positively correlated. Hence, both the QRF and the Eurosystem forecasts presents similar "deviations" from linearity, despite the fact that the Eurosystem modelling toolbox is overwhelmingly linear. This finding suggests that the judgemental component of the Eurosystem forecasts tends to embed a mild non-linearity on the projected inflation dynamics. Overall, it is remarkable that the QRF produces predictions which are competitive with those of the SPF and the Eurosystem forecasts, given that both of the latter sets of forecasts incorporate also expert judgement informed, among other things, by the news on likely future events, such as VAT increases, fiscal plans or the invasion of Ukraine.

This paper contributes to the very large literature on inflation forecasting. For a survey of the literature, see Faust and Wright (2013). In addition, our paper relates to a growing literature on the virtues of (machine learning) ensemble methods, which are becoming more and more popular in the econometric literature for prediction (Athey et al., 2019; Avramov, 2002; Bai and Ng, 2009; Clark et al., 2021; Faust et al., 2013; Fernandez et al., 2001; Inoue and Kilian, 2008; Jin et al., 2014; Ng, 2013; Rapach and Strauss, 2010; Sala-I-Martin et al., 2004; Varian, 2014; Wager and Athey, 2018; Wright, 2009). Giannone et al. (2021) shows that ensemble methods may be particularly successful thanks to their ability to appropriately handle model uncertainty. Medeiros et al. (2021) shows that the random forest helps to predict US inflation. We focus on the euro area and, more importantly, we focus on density forecasts, which are a crucial input for the risk assessment at the core of monetary policy decisions. The emphasis on density forecasts is still not at the center stage in the literature on machine learning, despite its relevance for policy-making institutions.

We also contribute to the literature on potential non-linearity in inflation dynamics (on the relevance of non-linearity for density forecasting, see Goulet Coulombe et al., 2022). Specifically, a large literature studies the likelihood of changes in the shape of the Phillips Curve and the factors which may potentially explain such changes. For an extensive survey and a systematization of the debate, see Del Negro et al. (2020). Several papers in this literature point to a different relationship of inflation with its determinants in high and low inflation regimes (see, for example Akerlof et al., 1996; Costain et al., 2022; Fahr and Smets, 2010; Benigno and Ricci, 2011; Lindé and Trabandt, 2019; Forbes et al., 2021; Clark et al., 2022).

The rest of the paper is organized as follows. In section 2, we discuss our empirical strategy. Section 3 presents the results of our out-of-sample forecasting accuracy assessment. Section 4 discusses the analysis of the contribution of different predictors to our inflation forecasts, based on the Shapley values. Section 5 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 of inflation at time t with its determinants at time t-h, for a generic forecasting horizon h. Then, we apply the estimated model on the data at time t to produce an inflation forecast at time t+h. The variable that we fit in our model is π_t^h, i.e. the annualized growth rate of the Harmonized Index of Consumer Prices or of the Harmonized Index of Consumer Prices excluding energy and food prices (HICP or HICPex, defined as P_t below), at the forecast horizons of 3, 6, 9 and 12 months ahead (h = 3, 6, 9 and 12):

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

Formally, we would like to estimate a non-linear relationship between our target concept of inflation π_t^h, its lags and a set of determinants x_{t-h}:

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

and then obtain an inflation forecast as

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

Rather than tightly parameterizing $m(.)$, we capture quite general forms of non-linearity by resorting to machine learning techniques. In particular, we estimate the potentially non-linear relationship of inflation with its determinants by means of the quantile regression forest (QRF) developed by Meinshausen (2006), which is a variant of the random forest of Breiman (2001), allowing for density forecasting.

A quantile regression forest is an ensemble method which combines the results from a certain (potentially large) number of non-linear models, called regression trees. A regression tree fits a specific target variable (headline or core inflation, in our case) by repeatedly splitting the sample of the potential predictors in different sub-samples. Once the final split is achieved, the predicted value of the target variable associated with a specific sub-sample is represented by the sample mean or median of the target variable in that sub-sample, called "leaf", for point prediction. In our paper, we focus on density prediction, which can be carried out by computing the empirical quantiles of the target variable associated with each leaf, as suggested in Meinshausen (2006). The sub-sample splits in a regression tree are obtained through a process defined as binary recursive partitioning, an iterative process that splits the data into partitions. The process continues until the splits achieve an improvement in terms of a statistical criterion, such as the mean squared error in the fit for inflation (our target variable) or, alternatively, until the splitting process hits a stopping rule which, in our case, is that any leaf contains at least ten data points. 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.[2]

2.2 Benchmark models

We compare the predictions from the quantile regression forest to several benchmarks.

First, we consider a state-of-the-art linear model, i.e. an equally weighted combination of 500 VAR models, which we define as VARCOMB. We choose as benchmark the combination of individual BVAR models to be as close as possible to the QRF, which is also a combination of models, i.e. regression trees. The main difference between VARCOMB and QRF lies in the possible nonlinearity captured by the latter. Each individual VAR model includes inflation (headline or core), plus four randomly selected indicators from our dataset. The data are stationarized, before entering the VAR models, and the latter are specified with two lags to mimic the procedures we follow in the QRF. The models are estimated using bayesian techniques. The prior distributions for the lag coefficients and error variances are in the Normal-Inverse Wishart class and are parameterized to shrink the model estimates toward the parameters of a random walk model, in the tradition of the Minnesota prior (Litterman, 1979; Doan et al., 1984; Banbura et al., 2010).[3] The prior hyperparameters are treated as random variables and their value is drawn from their posterior, following Giannone et al. (2015).

Our second benchmark consists in the headline and core inflation density forecasts from the ECB Survey of Professional Forecasters (SPF).[4] The SPF is conducted on a quarterly basis and its participants are experts affiliated with financial or non-financial institutions based within Europe. For this paper, we gathered the historical vintages, aggregated across experts, appearing at the beginning of February 2002 until November 2022, for headline inflation, and February 2017 to November 2022, for core inflation. The density forecasts are available for two definitions of

[2] 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.

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

[4] 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.

inflation, i) year-on-year inflation[5] and ii) inflation in the current year.[6] In terms of methodology, the experts use a mix of models and expert judgement and provide a probabilistic assessment of inflation falling in certain pre-specified ranges of values.

Our third benchmark consists in the Eurosystem headline and core inflation projections (BMPE, in short).[7] These institutional forecasts are also prepared on a quarterly basis and are published at the beginning of the third month of each quarter. For these forecasts, we look at the vintages published from March 2002 to December 2022, both for headline and core inflation. Notice that the BMPE do not provide a density forecast for a large part of the sample, so we only compare these projections to the point forecasts of the QRF. In terms of methodology, the BMPE are also based on model analysis complemented by expert judgement.

In our forecasting evaluation, we adapt the data availability of the QRF and the VARCOMB models to mimic the data availability of the SPF and the BMPE forecasters. Appendix A provides more details on how we match the timing across the different sets of forecasts.

For our assessment of the density forecasts, it is convenient to work with a probability density function. Therefore, for all our density forecasts (QRF, VARCOMB and SPF), we follow the practice in the literature (see, for example, Adrian et al., 2019) and fit a skew-t distribution (Azzalini and Capitanio, 2003).[8] 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.

In the appendix B, we also provide a comparison of the median QRF forecasts with the forecasts from a random walk model (RW), a popular benchmark of non-forecastability, which as in Atkeson and Ohanian (2001), forecasts inflation at time "t+h" as

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

2.3 Data

Beside headline HICP and HICP excluding energy and food, our two target variables, our database contains 60 variables. The data is obtained from the ECB Statistical Data Warehouse (SDW) and comes from a variety of original sources. Broadly speaking, the dataset is inspired by the

[5]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".

[6]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".

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

[8]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.

Phillips Curve framework, covering different areas of the economy, and the choice of the variables is similar 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 on prices, employment expectations, confidence measures, industrial production, euro area business cycle indicators, various productivity measures); measures of inflation expectations (for example, survey and market-based measures over different forecast horizons); 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

Our out-of-sample exercise is based on a recursive updating scheme. When implementing the recursive scheme, we align with the date of publication of the survey and institutional forecasts, which have a quarterly frequency. In other words, for the sake of the comparison with those benchmarks, we run the QRF and VARCOMB only once per quarter, and with a data availability that is comparable to that of the SPF and the BMPE forecasters.[9]

Specifically, first, we estimate our models with data up to December 2001 (which is our first "t"), as the SPF first cut-off date is around mid-January 2002. Instead, when comparing to the BMPE, we estimate using data up to January 2002. 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 quarter (effectively, three months) 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.[10]

The target variable for which we compute our measures of forecasting accuracy, both for headline and core HICP, is defined in terms of year-on-year growth rates. We adopt this convention because some of the benchmarks against which we compare (for example, the BMPE) are not seasonally adjusted. In other words, for the generic horizon h, we compute the measures of forecasting

[9]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.

[10]The evaluation of the QRF and VARCOMB forecasts, released at the monthly frequency, gives the same results as the evaluation based on the quarterly frequency and is available upon request.

accuracy in terms of the variable[11]

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

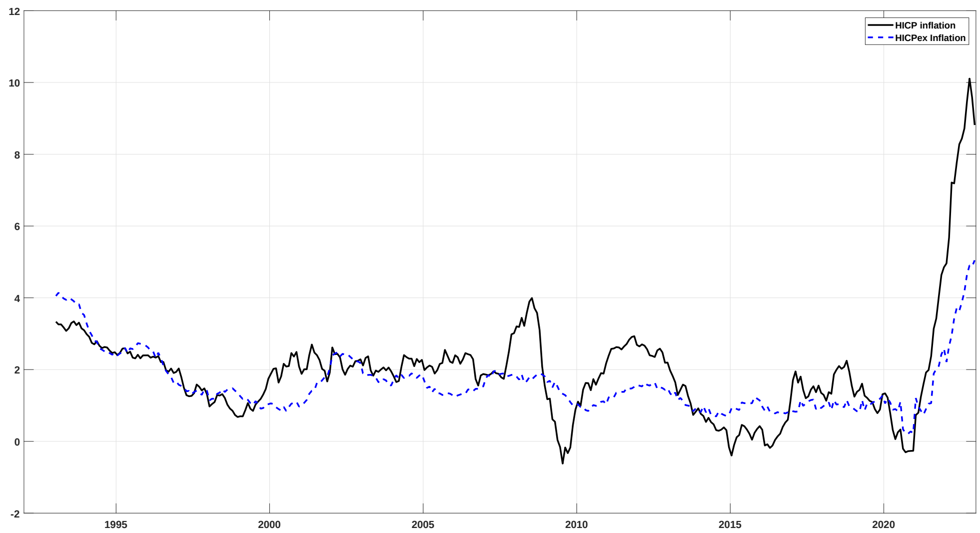
The only exception to this rule is for the exercise in which we compare our density forecasts with those of the SPF for the *current year*, where we conform to the practice of the SPF, which reports (headline and core) inflation in the current year in terms of the average year-on-year growth rate over the four quarters of the year.

In order to gauge the ability of the different models to capture the dynamics of inflation in different regimes, we focus both on the average accuracy over the whole sample and on the evolution of the measures of accuracy over time (Giacomini and Rossi, 2009, 2010; Rossi and Sekhposyan, 2016).

3 Results

Figure 1 reports the year-on-year growth rates of HICP (solid line) and HICPex (dashed line), our target measures of headline and core inflation.

Figure 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.

The figure shows the different regimes through which euro area inflation went over time. Notably, after the convergence toward the level of about 2% achieved in the early 90's, both headline

[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".

and core inflation were quite stable until the financial crisis of 2007-2009. In the run-up to the financial crisis, headline inflation markedly increased, fueled by a large increase in commodity prices, while core inflation remained stable. The recession ensuing from the financial crisis led to a sudden and large drop in headline inflation and a more delayed slowdown in core inflation. After the Great Recession and the initial rebound of inflation, headline and core entered a protracted period of relatively low inflation. Finally, the post-pandemic environment has been characterized by a sudden increase in both headline and core inflation to levels which, especially for headline inflation, are unprecedented in the euro area sample. Kuik et al. (2022) describes the role played by the turmoil in energy markets caused by the Russian invasion of Ukraine for the increase in euro area inflation. After the initial boost to inflation affecting mainly the energy component of inflation, the upside pressure on consumer prices became more broad based and, in the course of 2022 has affected the whole basket of prices, leading to a strong increase also of core inflation[12].

3.1 A comparison of the density forecasts of QRF versus state-of-the-art linear and judgemental forecasts

As a first step of our forecasting evaluation, we assess whether the density forecasts produced by the QRF, the VARCOMB and the SPF are correctly calibrated. A density forecast is correctly calibrated if, once 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 as $p(y_t)$ a generic density forecast, we assess its correct calibration by testing whether the probability integral transform (the cumulative density function corresponding to $p(y_t)$, PIT in short) of the realizations of the y_t process is distributed as an $U(0,1)$ (Diebold et al., 1998). Several methods to test for correct calibration have been proposed in the literature (see, for example Diebold et al., 1998; Berkowitz, 2001; Corradi and Swanson, 2006; Hong et al., 2007; Knüppel, 2015; González-Rivera and Sun, 2015). We rely on the test procedure described in Rossi and Sekhposyan (2019), based on the Kolmogorov-Smirnov test, which has also a graphical representation.[13] Figure 2 presents the results for headline inflation and Figure 3 for core inflation.

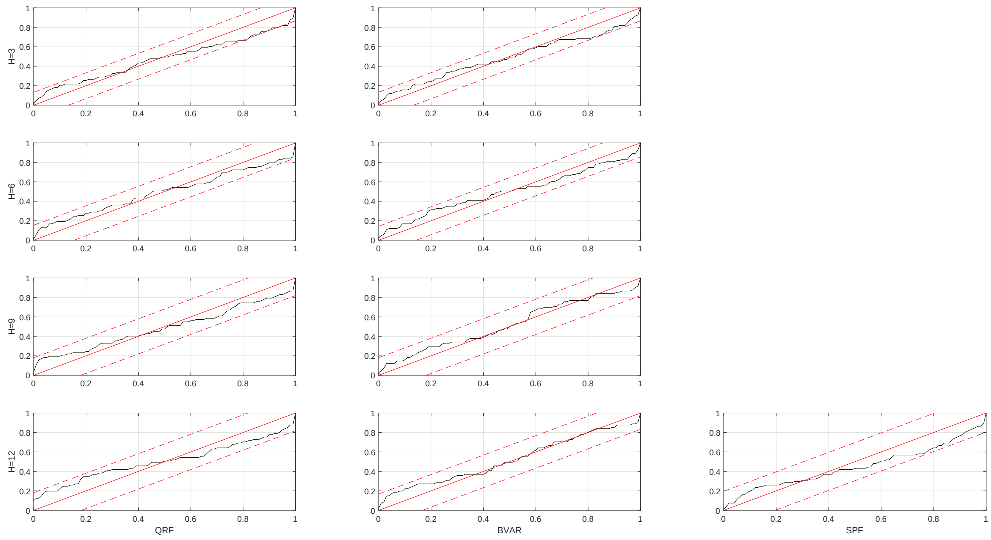
We accept the null hypothesis that the QRF, VARCOMB and SPF density forecasts for headline inflation are well calibrated. For core inflation, we have a slightly different picture. Both the QRF and VARCOMB[14] are less well calibrated, and for VARCOMB we also fail to accept the null of well calibrated forecasts for the forecasting horizons beyond three months. As in the empirical application of Rossi and Sekhposyan (2019), we find that the VARCOMB forecasts tend to be positively biased, especially in the last decade before the pandemic (see also figure 5).

[12] See Giannone et al. (2014) for a quantification of the pass-through of commodity price shocks to core inflation components

[13] 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).

[14] The SPF euro area core inflation forecasts are only available since 2017 and, hence the sample at our disposal is too short for a reliable assessment of the forecast accuracy.

Figure 2: Headline Inflation, test of uniformity of PITs

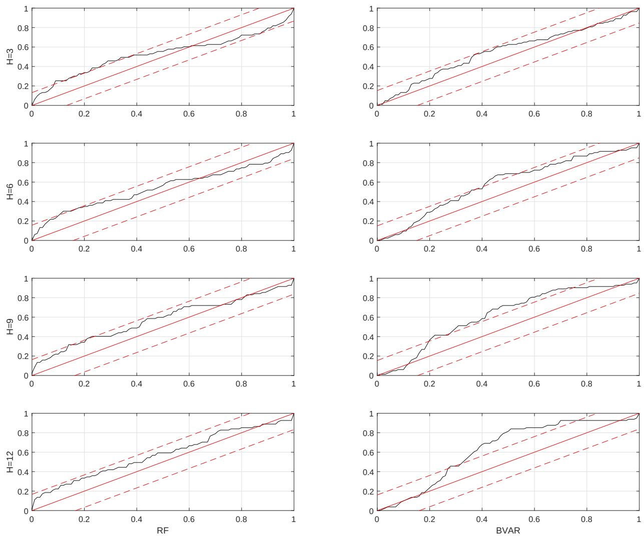


Note: Red lines: 1% critical values of the Kolmogorov-Smirnov test of PIT uniformity (dashed) and 45% degree line; Black line: Cumulative distribution function (CDF) of the PITs. If the probability density function of the PIT is a U(0,1), the CDF should be the 45% degree line. Left Column: QRF; Middle Column: VARCOMB; Right Column: SPF. The four rows correspond to the four forecasting horizons in the paper.

Calibration is a desirable property for density forecasts. However, Hamill (2000) highlights how calibration is only a necessary condition for a model to mirror the ideal forecaster, i.e. to perfectly capture the actual cumulative distribution function. Gneiting et al. (2007) argues that maximizing sharpness, given calibration, helps to better approximate the ideal forecaster. For this reason, we also evaluate the relative accuracy of the density forecasts by a proper scoring rule, i.e. the continuous ranked probability score (CRPS) of Gneiting and Raftery (2007). The CRPS measures the "distance" of the predictive cumulative distribution function from the empirical cumulative distribution function associated with the observations of the target variable. The lower the CRPS, the more accurate a specific density forecast. Scoring rules such as the CRPS measure simultaneously calibration and sharpness (i.e. the concentration) of density forecasts. Hence, looking at the CRPS allows us to complement the assessment of calibration conducted above. Another advantage of scoring rules is that they also allow us to rank different models. Table 1 reports the results for headline (Panel A) and core inflation (Panel B).

For headline inflation, we find that the QRF and VARCOMB have almost the same accuracy at the horizons of three and six months ahead. For the longer horizons, instead, the VARCOMB is more accurate than the QRF. For core inflation, instead, the QRF shows either a comparable or a slightly more accurate forecast accuracy than VARCOMB at all forecasting horizons. In general, our results suggest that the QRF is competitive with the state-of-the-art linear benchmark, particularly at the short horizons.

Figure 3: 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. Left Column: QRF; Middle Column: VARCOMB; Right Column: SPF. The four rows correspond to the four forecasting horizons in the paper.

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

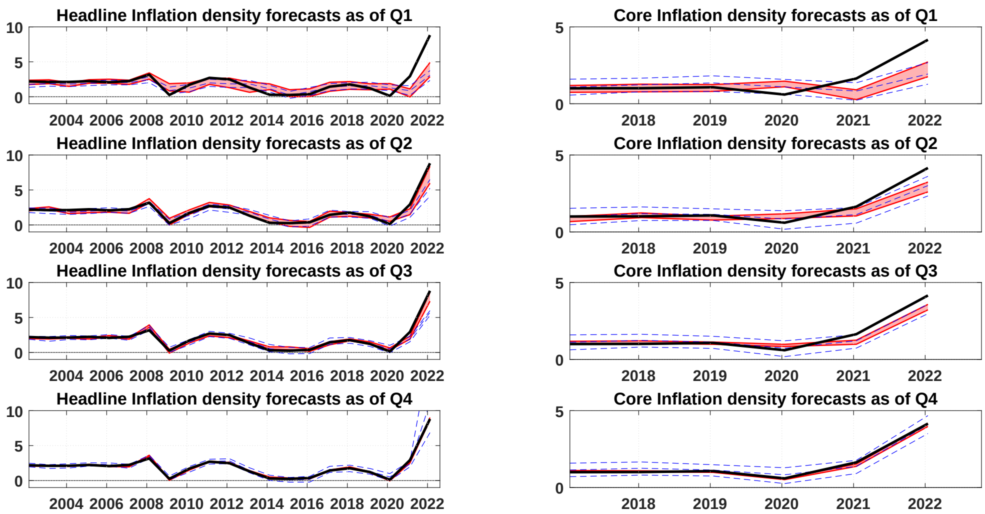
Horizon	**QRF**	**BVAR**	**SPF**
Panel a: Headline Inflation			
h=3	0.29	0.28	
h=6	0.50	0.49	
h=9	0.74	0.67	
h=12	0.93	0.88	0.87
Panel b: Core Inflation			
h=3	0.14	0.14	
h=6	0.23	0.24	
h=9	0.31	0.32	
h=12	0.37	0.39	

Note: CRPS for QRF (second column), VARCOMB (third column) and SPF (fourth column). The SPF are only available for the one-year-ahead forecasting horizon.

When we restrict our attention to the target defined in terms of year-on-year growth rates, the comparison with the SPF can only be carried out for the horizon of one year ahead and headline inflation. At that horizon, the accuracy of the SPF density forecasts is comparable to that of VARCOMB and it is superior to that of the QRF.

In order to get an idea of how the QRF compares to the SPF at shorter horizons than one year ahead, we assess the relative accuracy of the QRF and the SPF forecasts to predict *inflation over the current year*, as defined in sub-section 2.2. The prediction of *inflation over the current year* is reported in the SPF at each quarter of the year and, hence, it allows us to assess the SPF density forecasting accuracy at horizons which are shorter than one year ahead. Figure 4 below reports the charts with the observed average inflation (headline inflation, left and core inflation, right) for QRF and the SPF.

Figure 4: Headline and core inflation, density forecasts of QRF and SPF for the current year



Note: Red Area: 16th to 84th quantile of the QRF, current year for headline inflation (left panels) and core inflation (right panels); Dashed Lines: 16th to 84th quantile of the SPF, current year for headline inflation (left panels) and core inflation (right panels). The four rows correspond to the four quarters of each year in which the assessment is made.

For headline inflation, we have data over the 2002-2022 sample, while core inflation results are based on a shorter sample, ranging from 2017 to 2022. Interestingly, the QRF is as accurate or, especially in the third and fourth quarter of the year, more accurate than the SPF. In particular, QRF density forecasts are sharper around the actual value of inflation than the SPF. To appreciate the quantitative relevance of this point, Table 2 reports the CRPS for current year headline inflation forecasts of the QRF and the SPF.[15]

Clearly, the QRF is more accurate than the SPF for short horizon forecasts. While this assessment has some limitations, because the SPF is conducted in real-time and we use ex post revised data for the QRF, it should also be noticed that the SPF is a judgemental forecast and it can make use of valuable information about the future which is not embedded in the QRF information set. Hence, it is quite remarkable that the QRF has a comparable, if not better, forecasting accuracy than the SPF.

[15]For core inflation we have too few points, so we don't report CRPS.

Table 2: CRPS of QRF and SPF for current year headline inflation forecasts

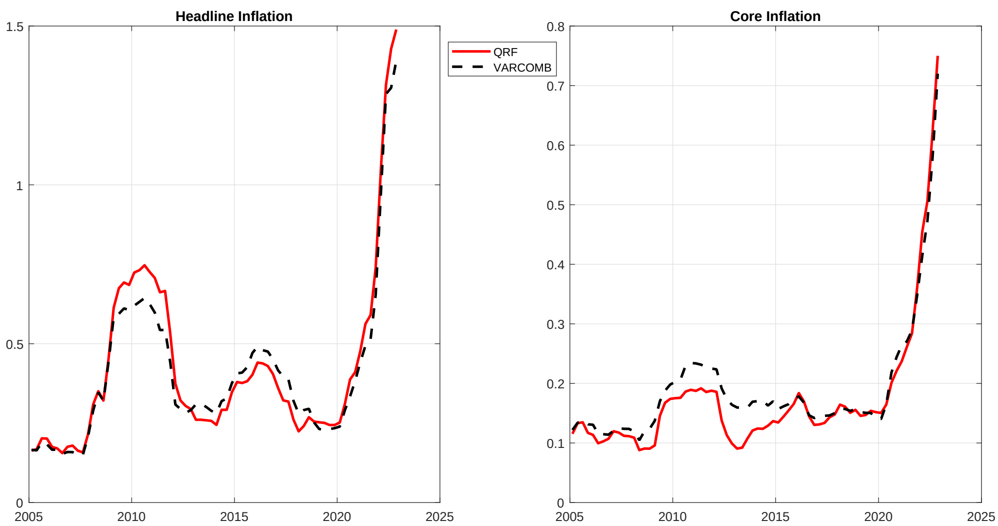
Quarter	QRF	SPF
Q1	0.56	0.61
Q2	0.25	0.33
Q3	0.18	0.29
Q4	0.08	0.21

Note: First column: quarter of the year in which the forecast is produced; Second column: CRPS of QRF; Third column: CRPS of SPF.

Overall, these results suggest that the QRF is a valid addition to the Eurosystem toolbox for inflation forecasting. Given the comparable accuracy with state-of-the-art linear models, the QRF is to be seen more as a complement rather than a substitute for the overwhelmingly linear Eurosystem forecasting toolbox. On the more general issue whether inflation dynamics are characterized by non-linearity, our results suggest that such non-linearity is probably not pervasive, but also that a mild non-linearity cannot be excluded, especially for core inflation.

In order to gauge the ability of the different models to capture the dynamics of inflation in different regimes (Giacomini and Rossi, 2009, 2010; Rossi and Sekhposyan, 2016) and, hopefully, to shed further light on the type of non-linearity in euro area inflation dynamics, figure 5 presents the CRPS of QRF and VARCOMB evaluated over rolling windows of three years. Again, we focus here on the horizon of six months ahead for both headline inflation (left panel) and core inflation (right panel), for brevity.

Figure 5: CRPS, three years rolling window, headline inflation at six months horizon



Note: Red solid line: QRF; Black dashed line: VARCOMB. The value on the vertical axis at each point refers to the average CRPS over the current quanrter and the previous eleven quarters.

Focusing first on headline inflation, VARCOMB is better able than QRF to account for the quick inflation rebound post Great Recession, detecting earlier the inflation trough and having been less reactive than the QRF throughout the crisis period. This result is in line with Ferrara et al. (2015) and Bobeica and Jarociński (2019) which show that linear models (with potentially large shocks) are able to accurately describe the inflation dynamics around the Great Recession. At the same time, the QRF adapts much faster than the VAR forecasts to the prolonged period of low inflation characterizing the pre-COVID decade. The accuracy of the non-linear model in this episode suggests that low inflation regimes may be characterized by different inflation dynamics than high inflation regimes, as hinted in Forbes et al. (2021), although our evaluation sample is relatively short, making the identification of high and low inflation regimes potentially challenging. Over the most recent sample, both the QRF and VARCOMB had some difficulty to capture the high inflation regime.

Interestingly, the right panel tells a different story for core inflation. Notably, for core inflation the QRF is superior to VARCOMB over most of the sample. The main difference between headline and core inflation is that the latter excludes the energy and the food prices from the HICP, two components which are obviously very affected by the dynamics of global commodity prices. Hence, considering together the results for headline and core inflation, our evidence suggests that the direct effects of commodity prices on headline inflation, especially via the energy components, are characterized by linear dynamics. When such direct effects of commodity prices are predominant for the dynamics of headline inflation, they dominate the non-linearity in the dynamics of the core inflation sub-component, making a linear model a competitive forecasting model for headline inflation in that regime.

3.2 Comparison of point forecasts with BMPE

The Eurosystem inflation forecasts (BMPE) have been reported as a point forecast for a large part of the sample under analysis and, hence, we will limit ourselves to a comparison of point forecasts. For the QRF, we consider the median of the density forecast distribution as point forecast. Table 3 reports the root mean squared errors (RMSE) for the QFR (left column) and the BMPE (right column), both for headline and core inflation.

The QRF point forecasts are generally comparable in accuracy to the BMPE forecasts at the short horizons and are less accurate at the nine and twelve month horizons. This result is remarkable, because the BMPE forecasts are the product of a very refined and sophisticated analysis by the Eurosystem forecasters and, via judgemental add-ons, they are flexible enough to embed all the available information on future events of an economic relevance, which may fail to be incorporated in the variables used in the QRF.

In figure 6 we plot the headline QRF inflation forecasts together with the BMPE and observed inflation for h=6.

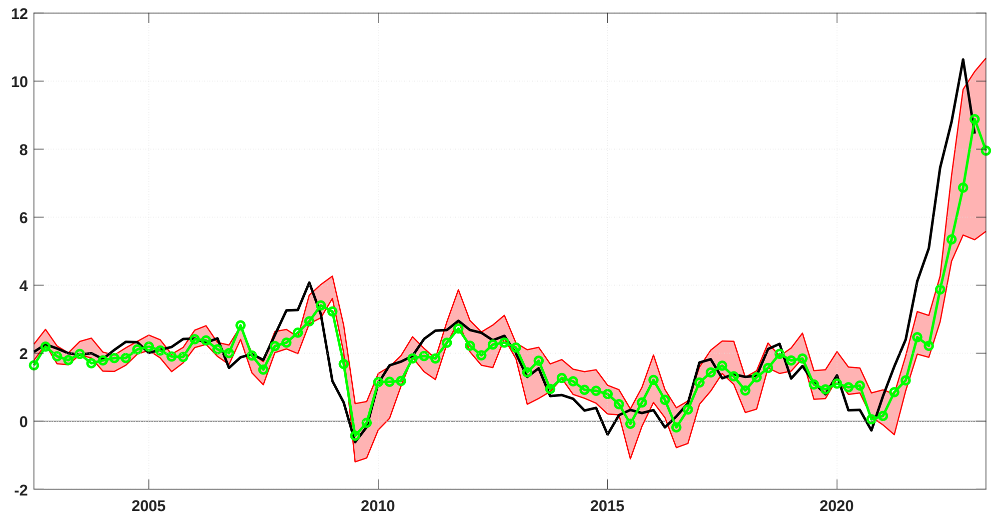
Even if the similarity between the two sets of forecasts is magnified by reporting year-on-year growth rates, the BMPE forecasts seem genuinely very collinear to the QRF forecasts. The

Table 3: RMSE of QRF and BMPE for headline and core inflation

Horizon	QRF	BMPE
Panel a: Headline Inflation		
h=3	0.58	0.47
h=6	0.92	0.94
h=9	1.48	1.42
h=12	1.97	1.65
Panel b: Core Inflation		
h=3	0.21	0.22
h=6	0.36	0.38
h=9	0.64	0.58
h=12	0.82	0.68

Note: Column 2: QRF; Column 3: BMPE.

Figure 6: Headline Inflation, density forecasts of QRF and BMPE, h=6

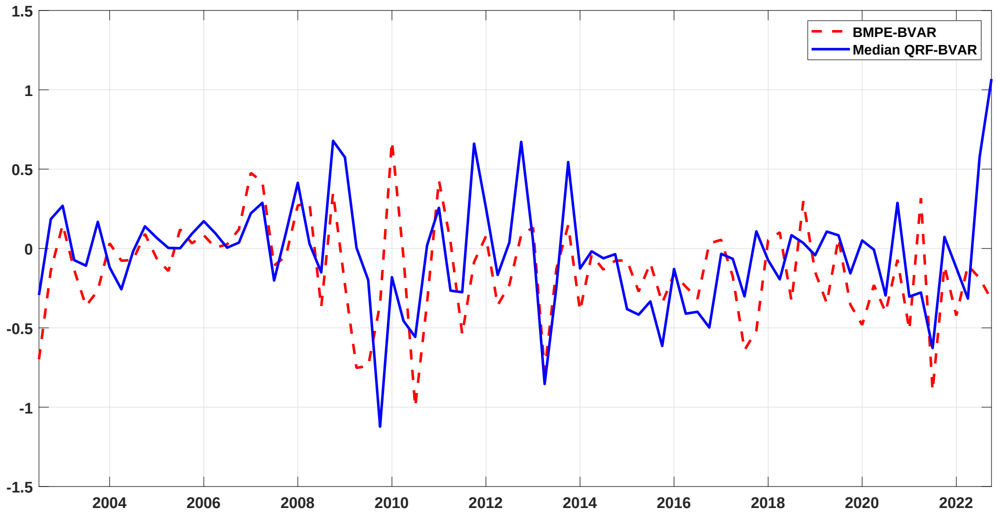


Note: Black solid line: year-on-year growth rate of HICP (headline inflation); Red area: 16th to 84th quantiles of the QRF density forecasts for the horizon of six months ahead, year on year growth rate of HICP; Green line with circles: BMPE projections for the horizon of six months ahead, year on year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

Eurosystem forecasts are produced by a toolkit that is essentially linear and, therefore, this result suggests that the BMPE judgemental component tends to introduce a non-linearity in the projections. Figure 7 shows the gaps of the median QRF forecast versus the median VARCOMB and the BMPE versus VARCOMB, which is a rough measure of "distance from linearity" of the two forecasts.

The two gaps are obviously correlated. Indeed, the correlation coefficient is about 0.4 for the three and the six month horizons and about 0.3 for the nine and twelve month horizons. This result

Figure 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.

suggests that judgement adds some element of mild non-linearity in the Eurosystem projections, rather consistently over time.

4 Shapley Values: interpretation of the forecasts and their functional forms

In this section, we study the drivers of our QRF predictions. For our analysis, we exploit recent advances in the machine learning literature (see Strumbelj and Kononenko, 2010; Lundberg and Lee, 2017; Buckmann and Joseph, 2022) which suggest to adopt the concept of Shapley values (Shapley, 1952) to define the contributions of the different variables to our predictions. In short, the Shapley value of a specific variable[16] for a forecast consists in the average marginal contribution of that variable to the forecast with respect to all the so called "coalitions" among variables, i.e. all possible combinations of variables in the predictor set. The marginal contribution of a specific variable to a coalition is the additional contribution from adding the variable to the coalition, once all variables not in the coalition have been integrated out. In the special case of a linear regression in which all the predictors are orthogonal with each other, the Shapley values of a variable are given by the regression coefficient times the deviation of that variable from its mean.[17]

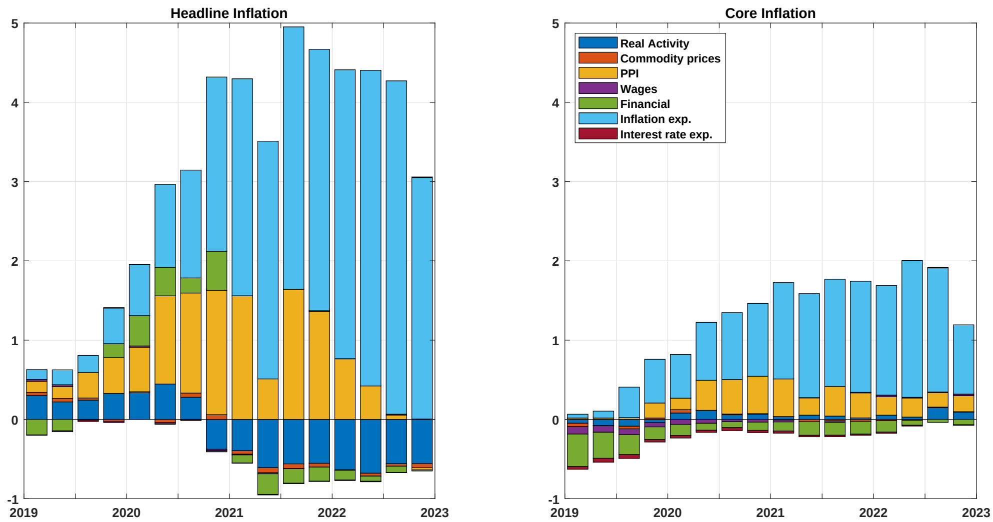
[16]In the literature on machine learning, the variables used as predictors are also defined as "features".
[17]See Aas et al. (2020) for a derivation of this result.

Generally, estimating Shapley values in presence of a large set of potentially correlated predictors is a rather complicated task, both for the large set of coalitions to be accounted for and the difficulties to correctly model the cross-correlation among variables when integrating out the effects of "off-coalition" variables. We rely on the method developed in Lundberg et al. (2019), which exploits the structure of regression trees underlying the QRF to speed up the computation of Shapley values and handle the issue of correlated predictors.[18]

We use the Shapley values for two main goals. First, we aim to define which predictors drive the inflation outlook. The predictors in our database are correlated among themselves, as it is generally the case for macroeconomic and financial variables. Hence, we do not aim to provide a fully fledged "narrative", i.e. a causal account of why our inflation predictions evolve in a specific direction, we only wish to study from which variables our model is extracting the signal for the inflation outlook.

As an illustration, figure 8 shows the results of the Shapley value analysis for headline (left panel) and core inflation (right panel), at the horizon of six months ahead, over the 2019-2022 period. The individual variables have been classified in groups (see appendix A) and their Shapley values have been aggregated to compute the contribution of the specific group to each prediction.[19]

Figure 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.

In the period 2019-2022, the QRF extracts the signal that inflation would be raising over the subsequent six months mostly from measures of inflation expectations and producer price indices.

[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.

The other variable groups have a negligible contribution to the forecasts. In particular, commodity prices do not seem to play a very large role for headline and core inflation. This is explained by the fact that some other measures (for example, inflation expectations) may have captured the signal that the boost in commodity prices would lead to higher inflation.

The Shapley values of individual variables are also useful to dig deeper on the question of which functional forms are captured by the QRF, suggesting whether inflation dynamics are characterized by non-linearity. Table 4 reports the top seven[20] individual contributors to the forecasts of headline and core inflation, at the horizon of six months ahead. The ranking is formulated on the basis of the mean absolute value of the contributions over the out-of-sample evaluation period.

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	Unemployment rate	Long-term interest rate future (6 months, DE)
5	Consumer survey - price trends over next 12 month	Unemployment rate
6	Ten-Year Govt Bond Yield	Inflation rate future (6 months, DE)
7	Industry survey - selling price expectations, Intermediate Goods	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.

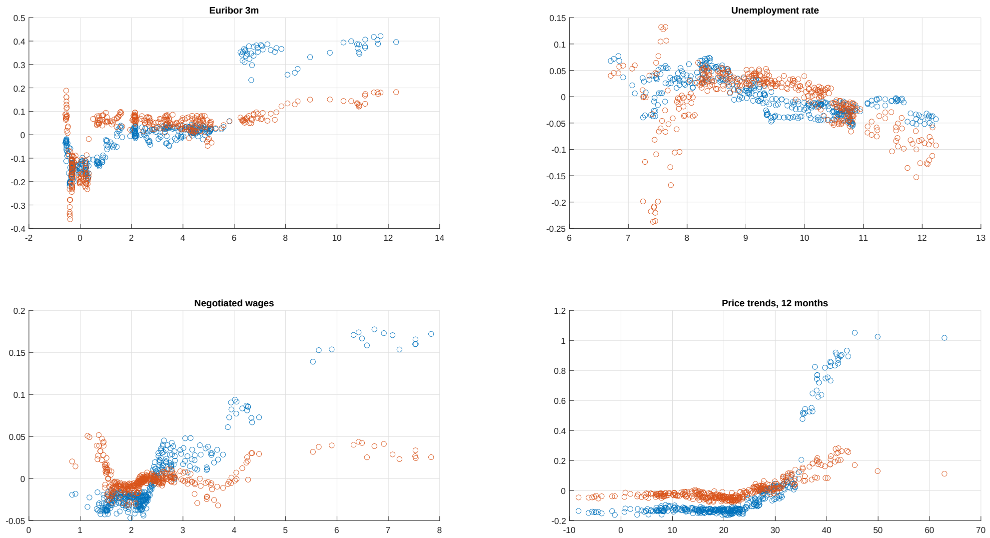
We find that short-term interest rates, measures of inflation expectations and real activity are important predictors of inflation. Which functional forms best characterize the relationship between these predictors and headline/core inflation? Figure 9 shows the Shapley values of some selected variables for headline (red circles) and core inflation (blue circles), plotted against the values of the variables over the historical sample. The variables are chosen among the most relevant indicators for inflation in the full sample, as reported in Table 4. These scatter plots give a rough indication of the type of the relationship that the QRF estimates between the variables and headline and core inflation for the six months forecasting horizon.[21]

For three of these variables, the relationship captured by the QRF seems roughly linear. For example, for the Euribor the relationship with inflation is essentially linear and, as expected, positive: high short-term interest rates signal that inflation is expected to be high in the future and needs some "leaning against the wind" from the central bank. The only non-linearity appearing in the Euribor charts has to do with the attainment of the effective lower bound of interest rates by the ECB. The panel on the unemployment rate effectively suggests the existence of a Phillips correlation, with a mild negative slope. The pass-through from wages to inflation is also roughly linear. For what concerns the measure of inflation expectations reported in the chart, i.e. "price trends expected over the next twelve months", the relationship with headline and, above all, core inflation is clearly non-linear. The measure of expectations is defined in terms of balance of survey respondents. A positive number indicates that there is a larger share of respondents who expect an increase rather than a decrease in prices. Hence, the figure suggests that the increase in the share of survey respondents who expect higher inflation beyond a certain threshold is an indication of a marked acceleration in inflation. The fact that the non-linearity in core inflation is

[20] Seven is roughly 10% of the variables in our set of predictors.

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

Figure 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

estimated to be more quantitatively relevant is in line with our finding in the previous section that the QRF is a better predictor for core rather than headline inflation, compared to VARCOMB, our state-of-the-art linear model.

5 Conclusion

In this paper, we show that the quantile regression forest, a non-linear model, could be a useful addition to the current Eurosystem toolbox to forecast euro area inflation, which is heavily skewed toward linear models.

The quantile regression forest has a comparable accuracy to state-of-the-art linear models, for density forecasting, and it is also competitive with institutional (BMPE) and survey (SPF) forecasts, despite the reliance of the latter on expert judgement.

At the same time, the similar accuracy of linear and non-linear models over the full sample under our analysis suggests the quantile regression forest is a complement rather than a substitute for the Eurosystem modeling toolbox to forecast inflation.

On the general question whether euro area inflation is characterized by non-linear dynamics, we conclude that non-linearity is mild and it is more evident for core than for headline inflation.

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 (HICPex). The former is what we consider our "headline" measure, while the latter is our "core" measure in this paper. 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).

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[22]. 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 (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 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 regularised mixtures of individual densities.

The first SPF vintage for headline inflation corresponds to February 2002, while the first one for core inflation is February 2017. The forecasts for those vintages are supposed to be produced aroundmid- 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.

[22] 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-MKT PRICES, RAW MATERIALS, EXCL.ENERGY(MU17), EUR-BASIS - HWWA . Euro area - HAMBURG WORLD ECONOMIC ARCHIVE
23	G2	2	WORLD-MKT PRICES, ENERGY RAW MATERIALS(MU17), EUR-BASIS - HWWA. Euro area - HAMBURG WORLD ECONOMIC ARCHIVE
24	G2	2	World market prices of raw materials, Index total, euro
25	G2	2	World market prices of raw materials, Index Total excluding energy, euro
26	G2	2	World market prices of raw materials, Energy, euro
27	G2	2	World market prices of raw materials, Crude oil, euro
28	G2	2	World market prices of raw materials, Industrial raw materials, euro
29	G2	2	World market prices of raw materials, Food and tropical beverages, euro
30	G2	2	ECB Commodity Price index Euro denominated, import weighted, Non-food
31	G2	2	ECB Commodity Price index Euro denominated, import weighted, Agricultural raw materials
32	G2	0	EXCH.RATE: US DOLLARS/1 EUR,SPOT AT 2:15 PM (CET) D,W,M,Q,A-AVG
33	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
34	G3	2	Producer Price Index, domestic sales, Consumer goods industry
35	G3	2	Producer Price Index, domestic sales, MIG Durable Consumer Goods Industry
36	G3	2	Producer Price Index, domestic sales, MIG Non-durable Consumer Goods Industry
37	G3	2	Producer Price Index, domestic sales, MIG Intermediate Goods Industry
38	G3	2	Producer Price Index, domestic sales, MIG Capital Goods Industry
39	G3	2	Producer Price Index, domestic sales, MIG Energy
40	G3	2	Producer Price Index, domestic sales, MANUFACTURING
41	G3	2	Producer Price Index, domestic sales, Total Industry (excluding construction)
42	G4	0	Indicator of negotiated wage rates, Total - Annual rate of change
43	G4	0	Indicator of negotiated wage rates - total excluding bonuses, Total - Annual rate of change
44	G1	2	Industrial Production Index, Total Industry (excluding construction)
45	G1	1	Building Permits / dwellings, Residential buildings except residences for communities
46	G1	0	European Labour Force Survey; Unemployment rate; Total; Age 15 to 74
47	G1	1	EA19 Leading Indicators OECD ¿ Leading indicators ¿ CLI ¿ Amplitude adjusted / Level. rate or national currency
48	G1	0	United States; European Labour Force Survey; Unemployment rate; Total; Age 15 to 74
49	G5	0	Euribor 3-month - Last trade price or value
50	G5	0	Benchmark bond - Euro area 10-year Government Benchmark bond yield - Yield
51	G5	2	European Monetary Union Market Index. Equity Index.
52	G5	0	IBES MSCI EMU Index Earnings. Weighted average long term growth EPS (Earnings per share) forecast expressed as a percentage
53	G5	2	Euro area - Equity/index - European Monetary Union Consumer Goods Index (EUR)
54	G5	2	Equity/index - European Monetary Union Consumer Services Index (EUR) - Historical close
55	G5	2	Monetary aggregate M1
56	G5	2	Monetary aggregate M3
57	G5	2	Monetary aggregate M2
58	G5	2	Loans, Total maturity, All currencies combined - Euro area (changing composition) counterpart
59	G6	2	US - CONSUMER PRICES, ALL ITEMS (ALL URBAN CONSUMERS)
60	G6	2	US - CONSUMER PRICES, CORE INFLATION (URBAN CONSUMERS)

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.

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 to our density forecasts

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. For the QRF, quantiles are directly available. For the SPF, we derive them from the SPF histograms.[23]

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 HICPex inflation and of the year-on-year growth rates of HICP and HICPex 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 5th/16th/84th/95th 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.[24] 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

[23]Notice that for VARCOMB we have already the draws from the posterior, which can be used to compute all our accuracy statistics. However, to produce the results in our tables, we also fit a skew-t to match the quantiles of the VARCOMB forecast. We do that for comparability purposes, but the results are basically unchanged if we use the original posterior draws.

[24]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.

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.

The skew-t distribution we obtain is then used to draw the density forecasts which enter of out-of-sample evaluation.

B Comparison of RMSE betweek median QRF and Random Walk forecasts

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

Horizon	**QRF**	**RW**
Panel a: Headline Inflation		
h=3	0.58	0.72
h=6	0.92	1.11
h=9	1.48	1.51
h=12	1.97	1.87
Panel b: Core Inflation		
h=3	0.21	0.31
h=6	0.36	0.45
h=9	0.64	0.61
h=12	0.82	0.75

Note: Column 2: QRF; Column 3: RW.

References

Aas, K., M. Jullum, and A. Løland (2020): "Explaining individual predictions when features are dependent: More accurate approximations to Shapley values," .

Adrian, T., N. Boyarchenko, and D. Giannone (2019): "Vulnerable Growth," *American Economic Review*, 109, 1263–1289.

Akerlof, G. A., W. R. Dickens, and G. L. Perry (1996): "The Macroeconomics of Low Inflation," *Brookings Papers on Economic Activity*, 27, 1–76.

Athey, S., M. Bayati, G. Imbens, and Z. Qu (2019): "Ensemble Methods for Causal Effects in Panel Data Settings," *AEA Papers and Proceedings*, 109, 65–70.

Atkeson, A. and L. E. Ohanian (2001): "Are Phillips curves useful for forecasting inflation?" *Quarterly Review*, 25, 2–11.

Avramov, D. (2002): "Stock return predictability and model uncertainty," *Journal of Financial Economics*, 64, 423–458.

Azzalini, A. and A. Capitanio (2003): "Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution," *Journal of the Royal Statistical Society, Series B*, 65, 367–389.

Bai, J. and S. Ng (2009): "Boosting diffusion indices," *Journal of Applied Econometrics*, 24, 607–629.

Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25, 71–92.

Benigno, P. and L. A. Ricci (2011): "The Inflation-Output Trade-Off with Downward Wage Rigidities," *American Economic Review*, 101, 1436–1466.

Berkowitz, J. (2001): "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business & Economic Statistics*, 19, 465–474.

Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013): "Time-varying combinations of predictive densities using nonlinear filtering," *Journal of Econometrics*, 177, 213–232.

Bobeica, E. and M. Jarociński (2019): "Missing Disinflation and Missing Inflation: A VAR Perspective," *International Journal of Central Banking*, 15, 199–232.

Breiman, L. (2001): "Random forests," *Machine learning*, 45, 5–32.

Buckmann, M. and A. Joseph (2022): "An interpretable machine learning workflow with an application to economic forecasting," Bank of England working papers 984, Bank of England.

Carriero, A., T. E. Clark, and M. Marcellino (2016): "Common Drifting Volatility in Large Bayesian VARs," *Journal of Business & Economic Statistics*, 34, 375–390.

Clark, T. E., F. Huber, G. Koop, and M. Marcellino (2022): "Forecasting US Inflation Using Bayesian Nonparametric Models," Working Papers 22-05, Federal Reserve Bank of Cleveland.

Clark, T. E., F. Huber, G. Koop, M. Marcellino, and M. Pfarrhofer (2021): "Tail Forecasting with Multivariate Bayesian Additive Regression Trees," Working Papers 21-08R, Federal Reserve Bank of Cleveland.

Conflitti, C., C. De Mol, and D. Giannone (2015): "Optimal combination of survey forecasts," *International Journal of Forecasting*, 31, 1096–1103.

Corradi, V. and N. R. Swanson (2006): "Predictive Density Evaluation," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Elsevier, vol. 1 of *Handbook of Economic Forecasting*, chap. 5, 197–284.

Costain, J., A. Nakov, and B. Petit (2022): "Flattening of the Phillips Curve with state-dependent prices and wages," *The Economic Journal*, 132, 546–581.

Darracq Pariès, M., A. Notarpietro, J. Kilponen, N. Papadopoulou, S. Zimic, P. Aldama, G. Langenus, L. J. Alvarez, M. Lemoine, and E. Angelini (2021): "Review of macroeconomic modelling in the Eurosystem: current practices and scope for improvement," Occasional Paper Series 267, European Central Bank.

de Bondt, G., E. Hahn, and Z. Zekaite (2018): "ALICE: A new inflation monitoring tool," Working Paper Series 2175, European Central Bank.

de Vincent-Humphreys, R., I. Dimitrova, E. Falck, and L. Henkel (2019): "Twenty years of the ECB Survey of Professional Forecasters," *Economic Bulletin Articles*, 1.

Del Negro, M., M. Lenza, G. Primiceri, and A. Tambalotti (2020): "What's up with the Phillips Curve?" *Brookings Papers on Economic Activity*, Spring.

Diebold, F. X., T. A. Gunther, and A. S. Tay (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

Diebold, F. X., M. Shin, and B. Zhang (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., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3, 1–100.

Ehrmann, M., G. Ferrucci, M. Lenza, and D. O'Brien (2018): "Measures of underlying inflation for the euro area," *Economic Bulletin Articles*, 4.

Elliott, G. and A. Timmermann (2016): *Economic Forecasting*, Princeton University Press.

Engelberg, J., C. F. Manski, and J. Williams (2009): "Comparing the point predictions and subjective probability distributions of professional forecasters," *Journal of Business & Economic Statistics*, 27, 30–41.

Fahr, S. and F. Smets (2010): "Downward Wage Rigidities and Optimal Monetary Policy in a Monetary Union," *Scandinavian Journal of Economics*, 112, 812–840.

Faust, J., S. Gilchrist, J. H. Wright, and E. Zakrajšsek (2013): "Credit Spreads as Predictors of Real-Time Economic Activity: A Bayesian Model-Averaging Approach," *The Review of Economics and Statistics*, 95, 1501–1519.

Faust, J. and J. H. Wright (2013): "Chapter 1 - Forecasting Inflation," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, 2–56.

Fernandez, C., E. Ley, and M. F. J. Steel (2001): "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, 16, 563–576.

Ferrara, L., M. Marcellino, and M. Mogliani (2015): "Macroeconomic forecasting during the Great Recession: The return of non-linearity?" *International Journal of Forecasting*, 31, 664–679.

Forbes, K. J., J. E. Gagnon, and C. G. Collins (2021): "Low inflation bends the Phillips curve around the world: Extended results," Working Paper Series WP21-15, Peterson Institute for International Economics.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann (2013): "Combining expert forecasts: Can anything beat the simple average?" *International Journal of Forecasting*, 29, 108–121.

Giacomini, R. and B. Rossi (2009): "Detecting and Predicting Forecast Breakdowns," *Review of Economic Studies*, 76, 669–705.

——— (2010): "Forecast comparisons in unstable environments," *Journal of Applied Econometrics*, 25, 595–620.

Giannone, D., M. Lenza, D. Momferatou, and L. Onorante (2014): "Short-term inflation projections: A Bayesian vector autoregressive approach," *International Journal of Forecasting*, 30, 635–644.

Giannone, D., M. Lenza, and G. E. Primiceri (2015): "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97, 436–451.

——— (2021): "Economic Predictions With Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437.

Gneiting, T., F. Balabdaoui, and A. E. Raftery (2007): "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society Series B*, 69, 243–268.

Gneiting, T. and A. E. Raftery (2007): "Strict proper scoring rules, prediction and estimation," *Journal of the American Statistical Association*, 102, 359–378.

González-Rivera, G. and Y. Sun (2015): "Generalized autocontours: Evaluation of multivariate density models," *International Journal of Forecasting*, 31, 799–814.

Goulet Coulombe, P., M. Leroux, D. Stevanovic, and S. Surprenant (2022): "How is machine learning useful for macroeconomic forecasting?" *Journal of Applied Econometrics*, 37, 920–964.

Hamill, T. (2000): "Interpretation Of Rank Histograms For Verifying Ensemble Forecasts," *Monthly Weather Review*, 129.

Hong, Y., H. Li, and F. Zhao (2007): "Can the random walk model be beaten in out-of-sample density forecasts? Evidence from intraday foreign exchange rates," *Journal of Econometrics*, 141, 736–776.

Inoue, A. (2001): "Testing For Distributional Change In Time Series," *Econometric Theory*, 17, 156–187.

Inoue, A. and L. Kilian (2008): "How Useful Is Bagging in Forecasting Economic Time Series? A Case Study of U.S. Consumer Price Inflation," *Journal of the American Statistical Association*, 103, 511–522.

Jin, S., L. Su, and A. Ullah (2014): “Robustify Financial Time Series Forecasting with Bagging,” *Econometric Reviews*, 33, 575–605.

Kenny, G., T. Kostka, and F. Masera (2013): “Can macroeconomists forecast risk? Event-based evidence from the euro area SPF,” *ECB Working Paper*.

Knüppel, M. (2015): “Evaluating the Calibration of Multi-Step-Ahead Density Forecasts Using Raw Moments,” *Journal of Business & Economic Statistics*, 33, 270–281.

Koester, G., E. Lis, C. Nickel, C. Osbat, and F. Smets (2021): “Understanding low inflation in the euro area from 2013 to 2019: cyclical and structural drivers,” Occasional Paper Series 280, European Central Bank.

Kuik, F., J. F. Adolfsen, E. M. Lis, and A. Meyler (2022): “Energy price developments in and out of the COVID-19 pandemic – from commodity prices to consumer prices,” *ECB Economic Bulletin*, Issue 4/2022.

Lenza, M. (2011): “Revisiting the information content of core inflation,” *Research Bulletin*, 14, 11–13.

Lindé, J. and M. Trabandt (2019): “Resolving the Missing Deflation Puzzle,” CEPR Discussion Papers 13690, C.E.P.R. Discussion Papers.

Litterman, R. B. (1979): “Techniques of forecasting using vector autoregressions,” Tech. rep.

Lundberg, S. M., G. G. Erion, and S.-I. Lee (2019): “Consistent Individualized Feature Attribution for Tree Ensembles,” .

Lundberg, S. M. and S.-I. Lee (2017): “A Unified Approach to Interpreting Model Predictions,” in *Advances in Neural Information Processing Systems*, ed. by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Curran Associates, Inc., vol. 30.

López-Salido, J. D. and F. Loria (2020): “Inflation at Risk,” Finance and Economics Discussion Series 2020-013, Board of Governors of the Federal Reserve System (U.S.).

Medeiros, M. C., G. F. R. Vasconcelos, Álvaro Veiga, and E. Zilberman (2021): “Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods,” *Journal of Business & Economic Statistics*, 39, 98–119.

Meinshausen, N. (2006): “Quantile Regression Forests,” *Journal of Machine Learning Research*, 7, 983–999.

Montes Galdon, C., J. Paredes, and E. Wolf (2022): “Conditional density forecasting: a tempered importance sampling approach,” *ECB Working Paper 2754*.

——— (2023): “A robust approach to tilting: parametric relative entropy,” Tech. rep.

Ng, S. (2013): “Variable Selection in Predictive Regressions,” in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. Granger, and A. Timmermann, Elsevier, vol. 2 of *Handbook of Economic Forecasting*, chap. 0, 752–789.

Probst, P., M. N. Wright, and A.-L. Boulesteix (2019): "Hyperparameters and tuning strategies for random forest," *WIREs Data Mining and Knowledge Discovery*, 9, e1301.

Rapach, D. and J. Strauss (2010): "Bagging or Combining (or Both)? An Analysis Based on Forecasting U.S. Employment Growth," *Econometric Reviews*, 29, 511–533.

Rossi, B. and T. Sekhposyan (2016): "Forecast Rationality Tests in the Presence of Instabilities, with Applications to Federal Reserve and Survey Forecasts," *Journal of Applied Econometrics*, 31, 507–532.

——— (2019): "Alternative tests for correct specification of conditional predictive densities," *Journal of Econometrics*, 208, 638–657.

Sala-I-Martin, X., G. Doppelhofer, and R. I. Miller (2004): "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94, 813–835.

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

Strumbelj, E. and I. Kononenko (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," *Journal of Economic Perspectives*, 28, 3–28.

Wager, S. and S. Athey (2018): "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests," *Journal of the American Statistical Association*, 113, 1228–1242.

Wright, J. H. (2009): "Forecasting US inflation by Bayesian model averaging," *Journal of Forecasting*, 28, 131–144.

Acknowledgements

The authors thank Joscha Beckmann, Laurent Ferrara, Francesco Furlanetto, Domenico Giannone and Giorgio Primiceri 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.

Michele Lenza
European Central Bank, Frankfurt am Main, Germany; Centre for Economic Policy Research, London, United Kingdom;
email: michele.lenza@ecb.europa.eu

Inès Moutachaker
Institut national de la statistique et des études économiques (INSEE), Paris, France; email: ines.moutachaker@outlook.fr

Joan Paredes
European Central Bank, Frankfurt am Main, Germany; email: joan.paredes@ecb.europa.eu

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

PDF ISBN 978-92-899-6115-8 ISSN 1725-2806 doi:10.2866/360772 QB-AR-23-067-EN-N