Forecasting Deflation Probability in the EA: A Combinatoric Approach *

Luca Brugnolini

European Central Bank and University of Rome "Tor Vergata"

[PRELIMINARY AND INCOMPLETE]

Abstract

I assess and forecast the probability of deflation in the EA at different horizons using a binomial probit model. I select the best predictors among more than one-hundred variables adopting a two-step combinatoric approach and exploiting parallel computation in Julia language. I show that the best-selected variables coincide to those standardly included in a small *New Keynesian* model. Also, I assess the goodness of the models using three different loss functions: the *Mean Absolute Error* (MAE), the *Root Mean Squared Error* (RMSE) and the *Area Under the Receiver Operating Characteristics* (AUROC). The results are reasonably consistent among the three criteria. Finally, I compute an index averaging the forecast to assess the probability of being in a deflation state in the next two years. The index shows that having inflation above the 2% level before March 2019 is extremely unlikely.

Keywords: inflation, forecasting, probability, Euro Area, ROC

JEL: C25, C63, E3, E58

---

* *The views expressed in this paper are those of the author and do not necessarily reflect those of the European Central Bank or the Eurosystem.*

Corresponding author: Luca Brugnolini luca.brugnolini@ecb.europa.eu, European Central Bank/Monetary Policy Strategy and PhD candidate at The University of Rome "Tor Vergata"

1 Introduction

Money predates history and price systems have been around as long as there has been money. In contrast, institutions with a stabilizing price mandate as central banks are newcomers, and as such, they continually need modern tools and new ideas to fill this gap. In fact, due to the complex nature of price changes, the achievement of the stability mandate is an arduous task, and central banks need to be very well-equipped to deliver. Although central banks can directly measure inflation with a regular frequency, the prototypical central bank toolbox is mainly composed by forecasting models; the reason is that central banks cannot rely only on contemporaneous inflation measures due to monetary lagged effects. In fact, monetary policy actions exerted today have a lagged impact which transmits to real variables solely in future periods. Against this background, forecasting is the leading alternative to ensure monetary policy to be timely and effective. Early assessing a possible inflation deviation from the target allows central banks to react. Such a complexity also explains why central banks employ an army of talented researchers and a tremendous quantity of resources in economic monitoring and forecasting: monetary policy is all about timing.

Given the extreme difficulty of their tasks, central banks have to revise and update their forecasting tools regularly. Among these, the most important are the models specifically developed to predict the inflation evolution. In the central bank parlance, the inflation evolution is commonly measured as the annualized percentage change in a consumer good and service market-based index, and it is critical relative to the central bank inflation target. The target is the benchmark to assess whether the prices of a country are stable. The theoretical reasons to keep inflation at a stabilized level comes to the economic theory. In fact, the inflation is one of the most informative variables about the stability of a country, and stabilizing inflation lays the foundations for the development of a modern economy, and ensures a moderate variation in other macroeconomic fundamentals. The last one is a well-known property, exposed in theoretical and empirical relationships as the *New Keynesian IS curve*, the *Phillips curve* and the *Taylor rule*. According to economic theory, through the *Taylor rule*, central banks can influence the real interest rate, and in turn output and inflation. Finally, an additional reason to desire price stability is that inflation is notably connected to financial markets and works as a background determinant of standard assets prices as sovereign bonds and currency.

In particularm within the Euro Area, the institution dealing with price stability mandate is the *European Central Bank* (ECB). For the ECB, the price stability objective is established by the *Governing Council* (GC) which is the decision-making body of the ECB itself. The original definition of price stability was an annualized increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below 2%. However, in 2003 the Governing Council has clarified that in reaching price stability it intends to maintain inflation rates *below, but close to, 2% over the medium term*. In achieving this target, the ECB rests on the investigation of the information based on the economic and monetary analysis[1], and among the tools employed, forecasting inflation is the one toward which more attention is devoted. For example, the ECB together with National Central Banks (NCBs) quarterly produces macroeconomic projections (MPE/BMPE). In particular, for forecasting inflation there are at least two different class of models; the first refers to fully micro-funded structural models as the *New Area-Wide Model* (NAWM) by Christoffel et al. (2008) and the *New Multi-Country Model* (NMCM) by Dieppe et al. (2011) and Dieppe et al. (2012). This class of models incorporates behavioral equations to model forward-looking economic agents to comply to the Lucas (1976) Critique. Also, these models incorporate short-term frictions mainly coming from the New-Keynesian school as price and wage stickiness. The advantage of this class of models, which often goes under the name of *Dynamic Stochastic General Equilibrium* (DSGE), is their ability to "tell stories" about the exogenous forces acting as drivers of the business cycle. However, the heavy structure imposed by model assumptions constrains the system dynamics and often produce unreliable results (Chari et al., 2009). Also, although DSGE are proven to be competitive in forecasting inflation (Faust and Wright, 2013), for computational reasons, their power is not yet exploited into full. In fact, DSGE models are typically log-linearized before being estimated and solved.

The second class of models employed by the ECB refers to time series models as *Vector Autoregressive* (VAR) models and their Bayesian counterpart (Giannone et al., 2014), and *Dynamic Factor* (DF) models. This class of models is endowed with less story-telling power. Nevertheless, they do not restrict the model parameters, which often results in much better forecast. However, there exist a third approach with which the ECB assesses the inflation evolution. This is by surveying more than 80 professional forecasters (so-called *Survey of Professional Forecaster* - SPF). The professional forecasters are specialists member of financial or non-financial institutions within the European Union[2]. In a typical survey, they are required to express their point forecast about inflation (as well as GDP growth and unemployment) over specific time horizons. Also, they are asked to provide probabilities for different inflation outcomes. For example, they are asked to give the probability that the year-on-year HICP inflation is below, in between or above certain thresholds. The

[1] This approach goes under the name of *two pillars strategy* and forms the information set under which the Governing Council takes its decisions.

[2] A detailed list of the participating organizations is available on the ECB website.

final forecast measure is the average of all the forecasts among forecaster. Although many papers highlight surveys' predictive ability (Faust and Wright, 2013), there are at least two main difference between them and proper in-house models which make the latters more attractive. First, there is an availability limit; SPF have a deterministic frequency in the release and cannot be updated as-soon-as there is the need. Secondly, there is an interpretation limit; in fact, the most informative measure coming from surveys is the median of the forecaster predictions, which summarize different models with different loss functions. Then, interpreting variation in the median is extrimely complex. Against this background, in this paper, I propose a third way between the forecasting model employed by the ECB and the probability measured in the SPF. In particular, I tailored a model to predict the inflation probabilities directly. Concerning the SPF, this tool has the advantage to be extremely easy to interpret given that the model specification is known. Thus, when a forecast shows some curious behavior, the forecaster can trace back the variable which causes it. Secondly, the model has the advantage to be updatable as soon as a new variable get released. While concerning DSGE, VAR and factor models, it has the advantage to be tailored for density forecast and not for point forecast. In fact, with continuous dependent variable models, a forecaster calibrates the model only on point forecasting, then compute the predictive density. In stead, with discrete model, a forecaster calibrate the model directly on density forecast. Secondly, using discrete models a forecaster can employ powerfull model selection criteria in addition to standard metrics as the *Mean Absolute Error* (MAE) and the *Root Mean Squared Error* (RMSE). The most prominent example is the *Area Under the Receiver Operating Charateristics* (AUROC).

In general, supporting point forecasts with probabilities provides a quantitative assessment of the forecaster uncertainty, however, probabilities are intrinsically informative also by themselves. In fact, knowing with which probability inflation over/undershoots the ECB target can help policymakers to decide on interest rate cuts or hikes. The argument can be heuristically formalized with the help of a Taylor rule as in Equation (1). The Taylor rule determines the central bank interest rate direction in response to price deviation from the target:

$$i_t = \phi_\pi(\pi_t - \pi_t^*) + \varepsilon_t \tag{1}$$

where i_t is the interest rate under the control of the central bank, π_t is the actual inflation rate and π_t^* is the inflation target. ε_t is a random variable defined in the literature as an unexpected monetary policy shock. The Taylor rule may also depend on other variables as the output-gap or lagged interest rate, but for an illustrative purpose, I abstract from these. According to Equation (1), when $\pi_t > \pi_t^*$

the central bank have to increase the interest rate while when $\pi_t < \pi_t^*$ the opposite has to happen. Following the rule, given ϕ_π, the point forecast is necessary and sufficient to know the magnitude of the interest rate adjustment. However, it is sufficient but not necessary to know the direction of the policymakers' actions. What is necessary is knowing whether inflation will be above or below the target and this information can be readily assessed through the probability to be in that range. Moreover, point forecast is more susceptible to forecast error than probability forecast, and accordingly, in this paper, I focus on buildin a tool to forecast the probability that inflation exceeds the central bank target.

To assess the probabilities of over/undershooting the inflation target, I investigate the predictive power of a large dataset of macroeconomic variables at different future horizons comprising macroeconomic and financial indicators and market surveys. Owing to dataset dimension, I perform a two-step variable selection procedure, and I use a combinatoric approach to retrieve the best model for each forecast horizon considered. The fundamental idea behind this process is that a valid forecast employs all the information available. Unfortunately, it is not always the case that standard models perform better including more data. The reason is in the risk of in-sample overfitting due to degrees of freedom depletion. In setting up the empirical exercise, I directly forecast probabilities using a binomial probit model. I choose the 2% inflation level as a natural cutoff point. In fact, many central banks have in their mandate this level as the inflation target, and the ECB approximately follow this rule as well (approximately because the exact objective is "below, but close to 2%"). In this respective, I forecast the probability of having inflation above/below the target at short and medium horizons.

In setting up this exercise, the main issue is that forecasting needs a precise horizons to predict. On the contrary, the correct implementation of monetary policy actions is related to a general medium-term orientation. The reason is that fluctuations in prices due to exogenous shocks make impossible to secure inflation at any point in time. Therefore, the intervals for achieving price stability has to be extremely general. The lack of a precise definition join to the delayed effect of the monetary policy actions makes the forecaster's life much more involved, imposing the need for a set of models calibrated for different horizons. Against this background, I propose a tool constructed by averaging the estimates of a set of forecasting models tailored for a grid of short to medium term horizons. The main idea connected to this choice is that macroeconomic as well as financial variables have different predictive power at distinct horizons, and a single model unlikely produces the best forecast at different steps-ahead. In this respective, I average the forecasted probabilities from the best horizon-calibrated models, and I create an index to predict the likelihood of having inflation above/below the 2% level in the next two years. The index shows that the probability of having inflation higher than 2% before March 2019 is extremely low.

2 Methodology

In this section, I describe the empirical methodology adopted in this paper. Firstly, I outline the process employed to build the dependent variable. In fact, the main difference concerning inflation probability forecast and standard recession prediction is in the choice of the dependent variable. Models tailored to predict the recession probabilities normally use as dependent variable a binary measure. This measure is computed by independent research organization which assesses and release a discrete variable to track recession periods [3]. For inflation, a clear counterpart does not exist. Nevertheless, a very satisfying and intuitive alternative can be found by clustering inflation realizations in points below and above the central bank target. Secondly, in this section, I outline the model evaluation procedure employed in this paper to select the best predictive variables at different horizons and in particular I describe the less-known AUROC metric.

2.1 Discretizing inflation

Within the Euro Area, many inflation metrics exist, however, all of them measure inflation as a continuous variable (π_t). In particular, the ECB definition of inflation target is in terms of year-on-year change in the *Harmonized Index of Consumer Prices* (HICP). There are other popular measures[4], however, as the paper focus on forecasting inflation from a central bank viewpoint and the ECB target is in terms of HICP, I will only focus on this measure.[5] To discretize the inflation measure and create the binary dependent variable (Π_t) I divide the HICP year-on-year change π_t into two different categories. I choose as a threshold the 2% level, as many central banks have this cutoff as a target, and I use it as an approximation for the ECB target. Thus our dependent variable looks as follows:

[3]For example, in the Euro Area the recession indicator is computed by the *Centre for Economic Policy Research* (CEPR), which is an independent organization. In the United States, the *National Bureau of Economic Research* (NBER) perform the same task.

[4]Fore example the *GDP deflator*, which is the ratio between nominal and real GDP or the *core inflation*, which is the HICP excluding food and energy

[5]Notice that forecasting from a central bank view point does not imply that only central banks can benefit from this study. Infact, spefic market agents, as speculators, can find more profitable to know the direction of inflation from the central bank viewpoint, more than knowing the value of the inflation itself. This is still because the central bank forecasts are the determinant of the central bank actions. In turn, actions, are powerful market movers if not anticipated.

- Inflation below the confidence zone ($\Pi_t = 0 \; if \; \pi_t < 2\%$).

- Inflation above the confidence zone ($\Pi_t = 1 \; if \; \pi_t \; \geq 2\%$).

The first panel of Figure 1 shows the yoy HICP for the EA (π_t) from January 1999 to March 2017. The solid blue line shows the monthly level in percentage points; the vertical grey bars highlight periods in which inflation is above or equal the 2% level, by contrast, the white bars shows periods in which HICP is below 2%). The second panel of Figure 1 shows the HICP sample distribution.

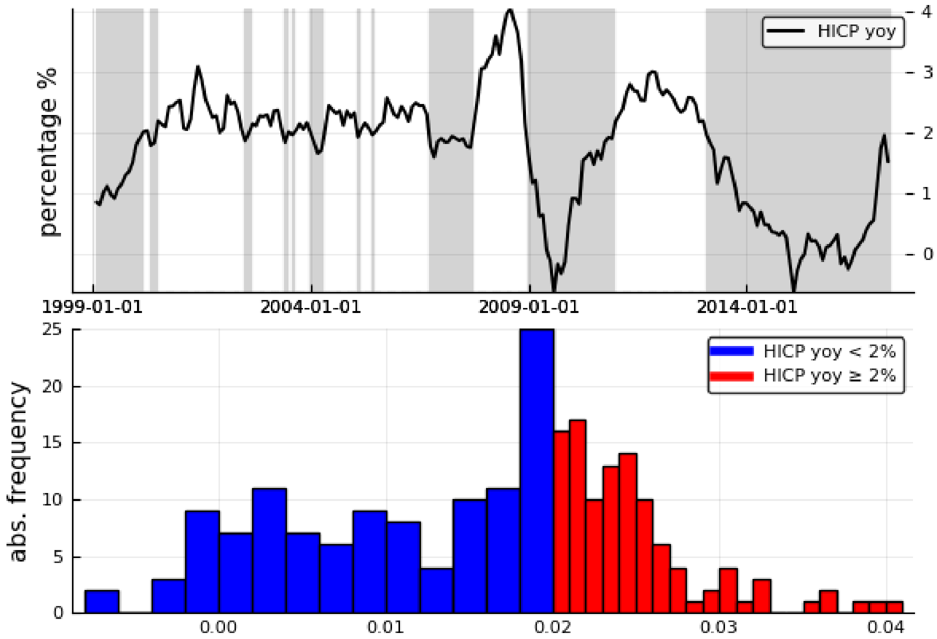

Figure 1: The blue solid line shows the yoy HICP for the EA, while bars show periods in which inflation is above or equal the 2% level.

The colors highlight the composition of the discretized HICP variables using 2% as a cutoff point. The blue bars (left-hand-side) show the portion of the distribution below the threshold while the red bars (right-hand-side) display the observations above or equal to it. The y-axis shows the absolute frequency of each bin. As built, the binary variable for inflation is well balanced along the entire sample. It displays 112 observation below the threshold and 107 above. From the chart, it is easy to notice that the mass tends to locate around the cutoff point. Indeed, the mode is located slightly below the 2% level, consistently with the ECB mandate. Also, it is interesting that while the right tail of the distribution concentrates around the threshold, the left tail is longer and exhibits more dispersion. This characteristic is mainly due to the recent deflationary period experienced by the Euro Area, which has led inflation in a negative territory for the first time after the great recession.

2.2 Model evaluation

To evaluate a model, the most important concept that a forecaster has to keep in mind is the loss function against evaluating the model itself. In the end, forecasting is a particular case of a decision theory problem (Elliott and Timmermann, 2016). In fact, different loss function reflects different weights a forecaster puts on the same forecast errors. And, of course, a different weighting scheme attributes to models with the same output different scores. This, in turn, affects model selection. Previous research has employed different metrics, but, the most selected are symmetric loss function as the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). The reason why researchers often select standard loss function is that, especially in economics, the choice of the loss function is often disregarded. This fact is true even if the selection of an appropriate loss function is intrinsically related to the problem the forecaster has to deal with. In fact, the only forecaster exempted from an accurate choice of the loss function is the one that commits no errors. Unfortunately, those who not belong to this class needs a metric to evaluate the distance of their prediction from the true realizations. The MAE and the RMSE are extremely valid loss function when a forecaster has symmetric disutility in over/underestimating the outcome, however, this is not always the case. Also, the MAE and the RMSE differ from the fact that the former penalizes errors more than the latter. Than, also models selected using this two standard loss function often select different variables. Equation (2) and (3) show the MAE and the RMSE using the variables described in this paper:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|\hat{\Pi}_{t+h|T} - \Pi_{t+h}\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\Pi}_{t+h|T} - \Pi_{t+h}\right)^2} \tag{3}$$

Where, $\hat{\Pi}_{t+h|T|}$ would be the predicted probability estimated from a discrete model. However, these particular loss functions are tailored for regression problems with continuous variables and built having in mind the evaluation of the model fit. In a discrete context, as the case of forecasting inflation probability, the prediction exercise resembles more a classification problem. Given the probability, I assign each forecast to one of the two class $\{0, 1\}$.

For this class of problems more appropriate loss function have been extensively studied in many scientific fields. Among these, a particularly well-tailored criterium for binary classification problems is the Receiving Operator Characteristics (ROC). The ROC is a metric computed in several steps, and in a nutshell, attribute a score to different models based on their ability to correctly classify observations among the whole spectrum of the possible cutoff points. The reader should notice that the cutoff here mentioned has nothing to do with the cutoff used to construct the dependent binary variable. In this context, the cutoff means the point in the estimated probability above which I classify an observation as a one or as a zero. In this sense, the main feature of the ROC is the ability to evaluate the categorization ability of a model along with all the possible cutoff and to attribute a score according to the best possible split. In this way, one does not have to evaluate models according to a particular threshold. Despite its proved ability, it has only been used in recent times in the economic literature (Berge and Jordá, 2011; Liu and Moench, 2016) and still it is unclear whether it provides a better assessment criteria than more standard methodology. I take a step in this direction by comparing in our exercise different models based on both the standard and the ROC methodology. In what follows I provide a brief description about how the ROC is computed.

The first step to compute the ROC is to evaluate the model ability to assign an observation to the correct class (*true positive/sensitivity, TP*) or to the wrong class (*false positive/fall-out, FP)* for all the possible threshold in the estimated probability. The threshold is a continuous variable approximated by a discrete variable constantly increasing by a tiny step between zero and one. Equation (4) and (5) shows the difference between the two types of thresholds.

$$\begin{cases} \Pi_t = 0 \quad if \quad \pi_t < 2\% \\ \Pi_t = 1 \quad if \quad \pi_t \geq 2\% \end{cases} \tag{4}$$

$$\begin{cases} \hat{\Pi}_t = 0 \quad if \quad \hat{\Pi}_t < C_i \\ \hat{\Pi}_t = 1 \quad if \quad \hat{\Pi}_t \geq C_i \end{cases} \tag{5}$$

Where $C_i \in [0,1]$, $i = 1, 2, ..., I$. For example, in the first step, a researcher estimates a discrete dependent variable model. In a discrete model, estimating the conditional expectation corresponds to estimate the entire model density. Therefore, the probability of having a particular outcome can be compared against a threshold C_i. Then the outcome can be classified accordingly. Repeating this process allows to assess the model classification ability. This is achieved by comparing the classification in (4) and (5). The result can be represented in a plane having the percentage of *TP* on the y-axis and the percentage of *FP* on the x-axis (*FP*(C_i),*TP*(C_i)). Figure 2 shows the ROC for a probit model estimated with the discrete version of inflation as the dependent variable and using as regressors the lag of HICP inflation.

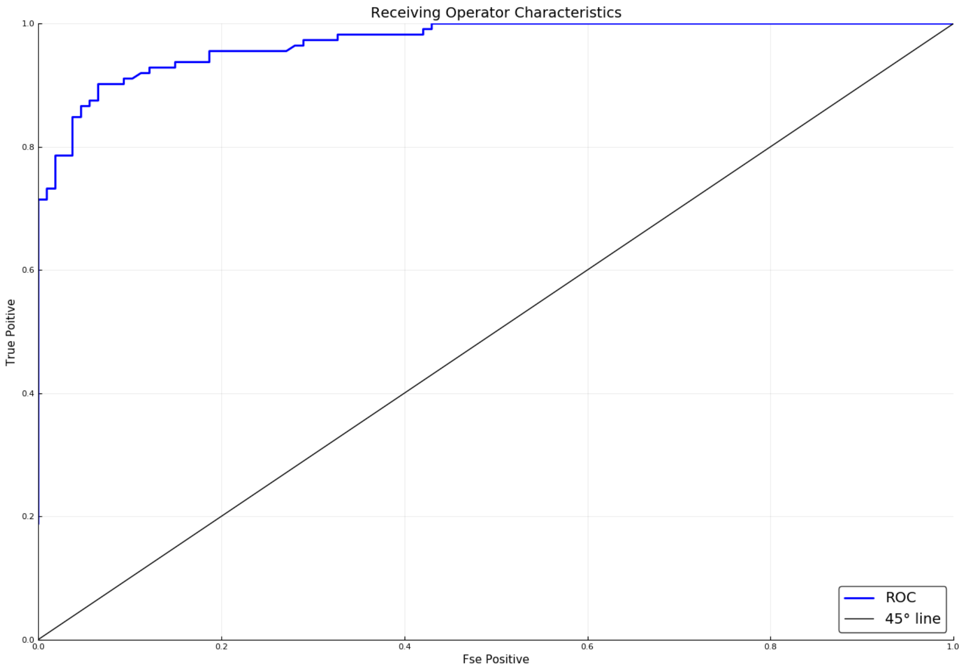

Figure 2: ROC curve computed with the estimated probability of having inflation above/below the 2% level by using lagged HICP and a constant in a univariate binomial probit model.

The ROC curve has the following interpretation; going from left to right, it gives the trade-off regarding false positives that one has to tolerate to increase the number of the true positives marginally. Other ROC characteristics are worth to be mentioned:

1. The 45-degree-lines has the equivalent meaning of a random guess (50% probability of having both *TP* and *FP*) and is often used as a reference line.

2. A ROC curve below the 45-means that one should revert the classification, to symmetrically flip the curve around the 45-lines.

3. Eventually, the best model attains 100% *TP* and 0% *FP,* which is the upper-left corner of the chart. This point gives the direction toward which the curve should increase to have a better model.

A scalar measure of the goodness of the model is the area under the ROC curve. As larger the area, as better the model. A commonly used estimator of the ROC area is the non parametric *AUROC* estimator shown in Equation (6):

$$AU\hat{R}OC = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left(X_i > Z_j + \frac{1}{2} (X_i = Z_j) \right), \quad AU\hat{R}OC \in [0.5, 1] \tag{6}$$

Where n_0, n_1 are respectively the number of zeros and ones according to the correct classification, X_i and Z_j are the estimated probabilities accordingly to the right classification. The $AU\hat{R}OC$ ranks models from the one with the largest area to the one with smallest.

3 Data

In this section, I briefly sketch the dataset used in the analysis. Also, I discuss the two-step model selection procedure developed to choose the most performing models.

3.1 Dataset description

I build a large dataset comprising around 100 monthly variables at national and Euro Area level starting in January 1999 and ending in March 2017 (219 observations). Table 9 in the Appendix shows the complete monthly dataset and the respective identification code. All the data comes from Thomson Reuters Eikon and Datastream except for the Wu and Xia (2016) shadow rate measure for the Euro Area as in Wu (2017) which are available on their webpage. The dataset has five different broad categories:

1. Real indicators
2. Price indicators
3. Monetary aggregates
4. Financial variables
5. Surveys

Each series is transformed to be approximately stationary. All the transformation code are reported in Table 9.

3.2 Model selection

The literature has developed many approaches to deal with large datasets. In general, these techniques either exploit dimensionality reduction such as factor models or employ parameter selection/shrinkage as the lasso estimator. However, the main drawbacks of these approaches are in terms of interpretability. Especially in forecasting, understanding which variables cause a change in the previsions is essential to judge the reliability of the forecast itself. For example, an experienced forecaster might guess that a variable is released higher than its consensus value or concerning other variables extremely correlated to that. An example here can help to clarify this point; suppose you want to forecast quarterly US GDP, and after the *non-farm-payroll* (NFP) release by the US Department of labor your forecasting model revises its prevision by a large number (let's say 0.5 percentage points). It could be the case that you are slightly puzzled by such a significant increase (even if there could be nothing wrong with that). Then, a valuable check can be looking to the value of the *ADP national employment report*, which is a variable always released a couple of days before the NFP and extremely correlated with the NFP itself. If you see a significant discrepancy between the two releases, you can put some probability on a measurement error that would be revised in next releases. Therefore, you may think about lowering your prediction by a value which depends on the discrepancy between the two variables weighted by the average difference between the two. This process is as plain as described only because of model interpretability. Unfortunately, this excellent characteristic gets lost in more sophisticated approaches. For these reasons, given that this paper focuses on forecasting the inflation probability, I build a direct subset selection approach based on two distinct steps which allow for the maximum possible degree of interpretability. Also, due to the prediction target, I focus on the out-of-sample performance of the model. The reason is that out-of-sample prediction and in in-sample fit are not always strictly linked. The best explanation for this is regarding the usual trade-off between bias and variance. In general, increasing the number of predictors in a model increases the in-sample fit by reducing the bias. However, this comes at the cost of overfitting the data, which often translates in a poor out-of-sample performance.

Having this statement in mind, in the present paper I focus exclusively on the out-of-sample performance of each of the model to evaluate their prediction performance. In particular, I develop a two steps selection procedure; in the first step, I perform a recursive out-of-sample analysis in a univariate regression framework. Meaning, for each variable in the dataset I fit a binomial probit model using Π_t as the dependent variable and that variable as the independent one. In selecting the forecast horizon, having in mind that the primary objective of the paper is an average prediction across short to medium period, I test each variable predictive power at eight different horizons. Being allowed to assess the predictive power of different variables at different horizons is an additional appealing feature to evaluate models using an out-of-sample metric. In fact, it is well-known among forecasters that different variables have different predictive power along different horizons. An exemplifying variable with this characteristics is the difference between the yield at long and short-term maturity, so-called yield curve slope (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998). The yield curve slope is renewed in the literature for being a powerful predictor of recessions. However, its predictive ability is evident only in the forecast at the medium-long horizons. On the contrary, it's short-term predicting strength is not as much good. Along with this line, I test different variables at different horizons to select a pool of predictors with a proved predictive power along different horizons. In what follows I describe the two main steps of the model selection procedure.

3.2.1 First step

In the first step, I regress $\Pi_{T+h|T}$ on each of the i variable in the dataset in a univariate probit model with a constant; in the notation $h \equiv [1, 3, 6, 9, 12, 15, 18, 24]$ is the forecast horizon which is an integer vector ranging between 1 to 24 months ahead and Π_{T+h} is conditioned on the information set available at time T. For each variable, I pre-estimate the model from January 1999 to March 2007 and compute a direct forecast up to the end of the sample. Equation 7 reports the model specification:

$$\Pi_{T+h|T} = G(x_t^{(i)}) + \epsilon_t, \quad h = [1, 3, 6, 9, 12, 15, 18, 24], \quad i = 1, ..., K \tag{7}$$

Where ϵ_t is an i.i.d normally distributed error term with variance restricted to $\sigma^2 = 1$. The normality assumption on the error term characterize the model as a probit model and attributes to the link function $G(.)$ the interpretation of the cumulative normal distribution function. [6]

Steps 1 to 4 describes the complete procedure:

1. For each variable i, pre-estimating the model coefficients on the period between 1999M1 to 2007M3.

2. For each horizon h, re-estimating the model recursively from 2007M3 to 2017M3 (120 months), increasing the sample size by one data-point for each iteration.

3. For each pair (i, h), assessing the model by comparing the differences between the estimated probability and the true probability using the *Area Under the Receiver Operating Charateristics* (AUROC), the *Mean Absolute Error* (MAE) and the *Root Mean Squared Error* (RMSE).[7]

4. Finally I select the two best variables for each horizon and each criterion.

Table 1 shows the best selected variables: Although it shows a heterogeneous pattern, it is possible to rationalize the results along with some common lines; first, it is evident that the best predictors

[6]For a review of the probit model check Wooldridge (2010) chapter 15.

[7]What I call true probability here is the probability of observing a certain outcome lying in a certain set after that the outcome is observed. For example, this implies that when $\pi_t = 1.5\%$ its probability of being in the set $\Pi_t < 2$ is one.

Table 1: Results from first step variable selection procedure

Horiz.	**AUROC**		**MAE**		**RMSE**	
$h = 1$	FR CPI SA	HICP FR	IT CPI SA	HICP IT	IT CPI SA	FR CPI SA
$h = 3$	FR CPI SA	HICP FR	IT CPI SA	HICP IT	FR CPI SA	IT CPI SA
$h = 6$	IP DE	EA7Y	EA7Y	EA3Y	EA7Y	EA3Y
$h = 9$	IP DE	Price trends 12M	EA7Y	EA5Y	EA7Y	EA5Y
$h = 12$	IP FR	Intermediate	EA7Y	EA10Y	EA10Y	EA7Y
$h = 15$	Intermediate	Industrial conf.	EA10Y	DE10Y	EA10Y	US10Y
$h = 18$	M3	Capital	US10Y	EA10Y	US10Y	EA10Y
$h = 24$	DE CPI SA	HICP DE	HICP DE	DE CPI SA	M1	DE CPI SA

Note: the table shows the best two predictors for each horizons selected among the entire dataset using a univariate probit model for forecasting the probability of having inflation below the 2% level. The prediction is evaluated according to three different criteria (AUROC, MAE, and RMSE) and the name of the selected variable is reported.

for very short horizons are some direct measures of inflation. Second, following MAE and RMSE, the best predictors for short-medium horizons are yields. In particular, the Euro Area interest rate between the 3 and 10-year maturity. It is interesting to notice that also the 10 year German and US government bond yield have some predictive power. This fact is likely due to the strong co-movements in the yield among markets. On the contrary, the AUROC selected predictors are more heterogeneous. From six months to one-year-ahead, the best predictors are real variables as the industrial production for Germany and France. However, also intermediate goods and capital for the Euro Area seems to have an outstanding predictive power, especially between twelve and eighteen-months-ahead. Also, it is fascinating to notice that beside real variables, also monetary variables as the M3 aggregate, surveys as the industrial confidence indicators and expectations of price-trends-in-one-year show an excellent forecasting power. Finally, for longer horizons, even if the monetary aggregate M1 shows some predictive power, the best predictors are some direct inflation measures as for shorter horizons. The interesting point to notice here is that for the shorter horizon, the best predictors were the inflation measures of France and Italy, while for longer horizons, the German inflation measure dominates. Understanding the economic reasoning behind these difference is out of the scope of this paper, but I suspect that the reason is intrinsically related to the weighting scheme used to build the EA HICP and to the disaggregated components of the national price index.

Table 2: Results from first step variable selection procedure. All criteria all horizons.

Price	**Interest rate**	**Real**	**Monetary**	**Survey**
HICP DE	EA3Y	Intermediate goods	M1	Industrial confidence
HICP FR	EA5Y	Capital	M3	Price trends 12M
HICP IT	EA7Y	IP FR		
DE CPI SA	EA10Y	IP DE		
FR CPI SA	DE10Y			
IT CPI SA	US10Y			

Note: the table shows the best predictors for all the horizons selected among the entire dataset using a univariate probit model for forecasting the probability of having inflation below the 2% level. The prediction is evaluated according to three different criteria (AUROC, MAE, and RMSE) and the name of the selected variable is reported. Each variable is reported only once, and it is allocated in one of the five macro-categories (Price, Interest rate, Real, Monetary, Survey)

Table 2 summarizes the 20 unique predictors delivered by the first selection step. Not surprisingly, the selected predictors coincide with variables considered the main determinants of inflation by established economic relationships. For illustrative purposes, consider a small-scale *New Keynesian model* as described by Equation (8) to (10) [8]. The first relation is the *New Keynesian IS curve* (NKIS) while the second describes the *New Keynesial Phillips Curve* (NKPC). The third relation is the Taylor rule introduced in the first section and reported here for convenience.

$$\hat{y}_t = \mathbb{E}_t\left[\hat{y}_{t+1}\right] + \frac{1}{\sigma}\left(i_t - \mathbb{E}_t\left[\pi_{t+1}\right]\right) + \epsilon_t \tag{8}$$

$$\pi_t = \beta \mathbb{E}_t\left[\pi_{t+1}\right] + \kappa \hat{y}_t + \eta_t \tag{9}$$

$$i_t = \phi_\pi(\pi_t - \pi_t^*) + \varepsilon_t \tag{10}$$

Where $\hat{y}_t$ is the *output-gap*, which is the difference between current output (y_t) and output at full employment (y_t^n), also called *potential output*. π_t is the inflation at time t, $\mathbb{E}_t\left[\pi_{t+1}\right]$ is the expected

[8] see for example Galí (2015).

inflation given the information set at time t, and i_t is a measure of monetary policy stance controlled by the central bank. ϵ_t and η_t are random variables referred to *technology* and *price mark-up shock*. For our purpose, the most relevant equation is the NKPC, which is a relation between current and expected inflation. However, if one belive in these simplified structure, selecting the best predictors among many variables should indeed return some proxy of the variables entering in the entire NK model. Table 2 shows exactly this point. First, the NKPC is a function of the expected inflation and the output gap. Among the selected variables, the expected price trend in the next twelve months can be considered a good proxy for the inflation expectation. Also, the NKPC links the price variation to the output-gap (which is a latent variable). For quarterly data, the best proxy for the output gap is the difference between current and potential GDP level. However, when a researcher deals with monthly variables, the best proxy for the output-gap is indeed constructed from industrial production. Indeed, from our variable selection procedure I retrieve exactly this variable for both France and Germany. Nevertheless, in the NK model, the output-gap is a function of the real interest rate, which is the difference between nominal interest rate and expected inflation. It is true that the interest rate included in NK models refers to the interest rate directly under the control of the central bank, however, due to strong comovements among yields, government bonds can be considered a good proxies for that variable. The selection procedure highlighted includes yields at different maturities among the best predictors for short to medium horizons. Finally, also the money supply which is often used in theoretical economic models as an alternative instrument under the control of the central bank to stabilize inflation, is reported among the best predictors.

From my point of view, this strong linkage between the variables selected in the first step and three established economic relationships build confidence in the procedure and benefit from the interpretability. Finally, it is valuable to notice that similar findings are common in the inflation forecasting literature. In fact, various forms of the NKPC are often estimated and used as a proper reduced-form model to forecast inflation directly. In the second step, I employ the twenty selected variables as an input for the model selection procedure as described in the next paragraph.

3.2.2 Second step

In the second step, I perform a procedure similar to *best subset selection*. I fit a separate probit model to all the possible combinations of the $K_2 = 20$ predictors selected in the first stage. Having twenty different variables implies that the number of possible combination is extremely high. Then, I restrict the number of maximum regressors in the model to $K_1 = 10$. This is the set which contains the largest number of combinations. Equation (11) shows the total number of possible combinations:

$$C = \sum_{k=1}^{K_1} \binom{K_2}{k} = 2^{K_2} - \sum_{k=K_1+1}^{K_2} \binom{K_2}{k} - 1 \tag{11}$$

As $K_1 = 10$ and $K_2 = 20$ in our setting, this lead to $C = 616,665$ models to estimate. Moreover, given the recursive structure of the out-of-sample exercise, this number has to be multiplied first by the number of data points by which the model is re-estimated. Those are $T^{out} = 121$. Secondly, it has to be multiplied times the number of horizons in the direct forecasting procedure for which I re-estimated the models, which are $H = 8$ and specifically $h = [1, 3, 6, 9, 12, 15, 18, 24]$. This process lead to estimate around a number of models $M \approx 600,000,000$. Finally, for each estimated model I compute a counterpart model augmented with a common factor extracted from the complete dataset. The common factor is estimated nonparametrically by computing the principal component from the eigenvector corresponding to the largest eigenvalue of the demeaned variables variance-covariance matrix. The reason to have an augmented counterpart models is that many authors, starting from Stock and Watson (2002) have proven to have very satisfactory performance in forecasting while providing a parsimonious way of dealing with large dataset. Stock and Watson (2016) surveys this class of models and their use in macroeconomics. Introducing this twist, from one side helps to explore important information left out of our process due to the presence of strong predictors. However, this heavily increases the estimation burden in terms of computation. In fact, first, adding a factor-augmented counterpart for each model double the models to estimate (which number approximately reach 1.2bn). Secondly, the principal component is recursively re-estimated when a data point is added, to avoid including information from the future. This process adds additional computational burden to our problem. To deal with such complexity, I write the entire code in Julia Language (Bezanson et al., 2017), and I perform estimation parallelizing the code on an octa-core processor. Julia is a modern and flexible open source language, which easily allow to perform parallel computing and to deal with computationally intense problems. Table 3 shows the time employed for the combinations of each variable groups, including the time for the out-of-sample performance and the principal component analysis.

I evaluate each model for each horizon according to the AUROC, MAE, and RMSE. Also, to have a benchmark for the comparison, I build a *naive model* fitted only the first lag of HICP for the EA and compare the performance of each model against this one. Finally, I select only models with maximum AUROC and minimum MAE and RMSE or each horizon. The second stage of the selection process returns a set of $3H = 24$ models.

Table 3: Computation time for each combination of variables in the second step.

#Variables	**#Combinations**	**Time**
1	20	13s
2	190	117s
3	1,140	12m
4	4,845	48m
5	15,504	2h34m
6	38,760	6h28m
7	77,520	13h
8	125,970	19h20m
9	167,960	28h30m
10	184,756	31h20m
Total	616,665	$\approx$ 100h

*Note: the table shows the number of variables used as regressors in a multivariate probit model (univariate when **#Variable** is equal to one) for forecasting the probability of having inflation below the 2% level. The total number of variables used are twenty and were selected in a univariate framework in a previous step. When the **#Variable** is equal to K, there are $\binom{20}{K}$ possible combination. The table also shows the amount of time employed by the Julia code for each particular number of combinations.*

Table 4: Results from second step variable selection procedure. AUROC criteria all horizons.

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 15$	$h = 18$	$h = 24$
AUROC	1.02	1.05	1.19	1.63	1.64	1.81	1.54	1.51
Factor	0	1	1	1	1	0	1	0
#Var	5	9	8	10	8	7	8	8
	Capital	M1	Capital	Inter.	Inter.	Ind.Conf.	Capital	M3
	Ind.Conf.	M3	M3	Capital	Capital	M1	M1	HICP DE
	IP FR	IP DE	IP DE	Ind.Conf.	Ind. Conf.	M3	M3	HICP FR
	IT CPI SA	IP FR	IP FR	M3	M3	EA10Y	IP DE	HICP IT
	FR CPI SA	US10Y	HICP IT	HICP DE	IP FR	IT CPI SA	IP FR	DE10YT
		DE10YT	EA7Y	US10Y	US10Y	FR CPI SA	US10Y	EA3Y
		EA3Y	DE CPI SA	DE10YT	DE10Y	PRICE 12M	EA10Y	EA5Y
		EA7Y	PRICE 12M	EA10Y	EA10Y		FR CPI SA	FR CPI SA
		EA10Y		DE CPI SA				
				PRICE 12M				

Note: the table shows the best model for each horizon h, selected among $\sum_{k=1}^{10}\binom{20}{k}$ used as regressors in a multivariate probit model for forecasting the probability of having inflation below the 2% level according to the AUROC criteria. The table also highlights the number of variables in the model and whether the forecast improves including the first principal component of the original dataset (Factor equal to one implies that including the factor enhances the prediction).

Table 5: Results from second step variable selection procedure. MAE criteria all horizons.

	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 15$	$h = 18$	$h = 24$
MAE	0.25	0.39	0.17	0.11	0.09	0.15	0.12	0.02
Factor	1	1	1	1	1	1	1	1
#Var	10	10	8	10	10	10	8	10
	Inter.	Inter.	Capital	Inter.	Inter.	Inter.	Capital	Capital
	Capital	Capital	M3	M1	Ind.Conf.	Ind.Conf.	M1	M1
	M3	Ind.Conf.	IP DE	M3	M3	M1	M3	M3
	HICP DE	M3	IP FR	IP FR	IP DE	M3	IP DE	IP DE
	DE10YT	HICP FR	HICP IT	HICP DE	IP FR	HICP DE	IP FR	IP FR
	EA3Y	HICP IT	EA7Y	HICP FR	HICP DE	HICP FR	US10Y	EA5Y
	EA5Y	EA3Y	DE CPI SA	HICP IT	DE10YT	US10Y	EA10Y	EA7Y
	EA7Y	EA5Y	PRICE 12M	EA3Y	IT CPI SA	EA7Y	FR CPI SA	EA10Y
	IT CPI SA	EA7Y		EA7Y	DE CPI SA	EA10Y		DE CPI SA
	DE CPI SA	DE CPI SA		PRICE 12M	FR CPI SA	FR CPI SA		FR CPI SA

Note: the table shows the best model for each horizon h, selected among $\sum_{k=1}^{10}\binom{20}{k}$ used as regressors in a multivariate probit model for forecasting the probability of having inflation below the 2% level according to the MAE criteria. The table also highlights the number of variables in the model and whether the forecast improves including the first principal component of the original dataset (Factor equal to one implies that including the factor enhances the prediction).

Table 6: Results from second step variable selection procedure. RMSE criteria all horizons. Factor equal to one means that the model include a factor estimated on the whole dataset using principal component.

	$h=1$	$h=3$	$h=6$	$h=9$	$h=12$	$h=15$	$h=18$	$h=24$
RMSE	0.64	0.75	0.35	0.31	0.37	0.39	0.21	0.16
Factor	0	0	1	1	1	0	1	1
#Var	8	4	8	8	10	6	8	10
	Ind.Conf.	Capital	Capital	Inter.	Inter.	Capital	Capital	Capital
	M1	Ind.Conf.	M3	M1	Ind.Conf.	IP DE	M1	M1
	M3	EA7Y	IP DE	M3	M3	IP FR	M3	M3
	IP FR	FR CPI SA	IP FR	HICP IT	IP DE	HICP DE	IP DE	IP DE
	US10Y		HICP IT	DE10YT	IP FR	EA10Y	IP FR	IP FR
	EA5Y		EA7Y	EA10Y	HICP DE	PRICE 12M	US10Y	EA5Y
	EA10Y		DE CPI SA	FR CPI SA	DE10YT		EA10Y	EA7Y
	FR CPI SA		PRICE 12M	PRICE 12M	IT CPI SA		FR CPI SA	EA10Y
					DE CPI SA			DE CPI SA
					FR CPI SA			FR CPI SA

Note: the table shows the best model for each horizon h, selected among $\sum_{k=1}^{10}\binom{20}{k}$ used as regressors in a multivariate probit model for forecasting the probability of having inflation below the 2% level according to the RMSE criteria. The table also highlights the number of variables in the model and whether the forecast improves including the first principal component of the original dataset (Factor equal to one implies that including the factor enhances the prediction).

The best-selected models are reported in Table 4 to 6. Depending on the criterion chosen, the results are mostly heterogeneous, both in the number of selected variables and in the inclusion of a common factor. Also, according to different criteria the difference between the naive model and the selected model is weaker or stronger. However, some common characteristics are worth to highlight. First, according to all criteria, the selected models are always able to outperform the naive model. However, for shorter horizons, the naive model is more difficult to beat. For longer horizons, the selected models perform much better. Secondly, for some specific horizons, the three criteria agree on both the number and the variables to include. Two clear example are the horizons $h = 6$ and $h = 18$. Thirdly, The AUROC and the RMSE are more parsimonious criteria in terms of the number of selected variables, while the MAE is the less. Forty, on average, it seems that all the models use predictors coming from different classes, implying that those can bring distinctive information useful to improving the prediction. Figure 3 shows the ratio between each of the selected models against the naive model. As expected, the naive model is performing better in the very short horizons. For $h = 1$ the best model selected according to the AUROC is only slightly better. However, since $h = 3$ the models selected using different criteria considerably outperform the naive model and present massive increases until $h = 9$. Then, depending on the criteria, the curve is stable or slightly

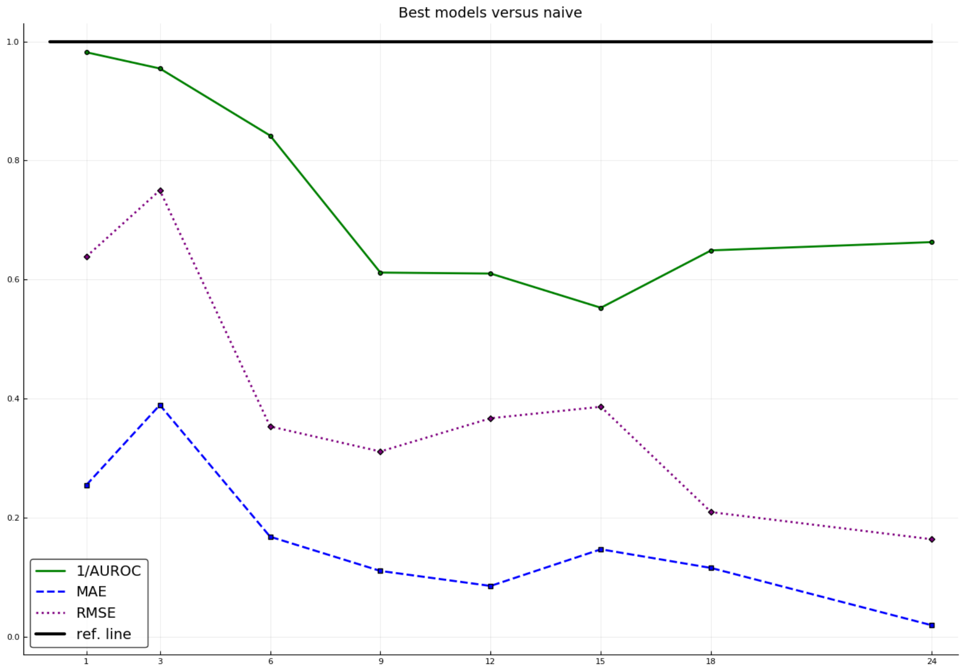

Figure 3: Ratio of the best selected model AUROC, MAE and RMSE over the naive model for each horizons. The black solid line highlights the sets of points for which the performance of the two models are equal.

increasing/decreasing.

4 Results

In this section, I analyze the in-sample and out-of-sample performance of the models selected in the previous section. The in-sample performance gives us a general framework to assess the overall performance of our model. The reason is that, given the lack of a long time series for the EA, there are only a few separate periods in the out-of-sample exercise in which inflation is below the 2% level. However, given that I am mainly interested in forecasting, after assessing the in-sample performance, I evaluate the out-of-sample performance of the selected models, and I predict in a true out-of-sample the probability of having low inflation from the end of the sample to March 2019. Finally, I build an index averaging all the forecast from one month to two years and assess the index prediction for the following years.

4.1 In-sample analysis

I start the analysis of the results of the selected models by assessing their in-sample fitting ability. I focus mainly on the probability that inflation is below the 2% level. The reason is that the downside risk seems the major concern in the EA. However, given that I am forecasting the whole density of the process, the upside risk can always be computed as the complement of the downside probability. Figure 4 shows the in-sample fit of the models. For each horizon, I plot the three best model selected according to AUROC, MAE, and RMSE. The grey bars represent periods of inflation is smaller then 2%. The chart shows an overall satisfactory in-sample fit. In particular, it is possible to notice that the first part of the sample is characterized by inflation above the threshold, except for some very small parenthesis. This behavior creates many false signals in the estimated probabilities, especially at shorter horizons. Starting from 2008 the series is characterized by prolonged periods of inflation above/below the threshold, which substantially reduces false signals.

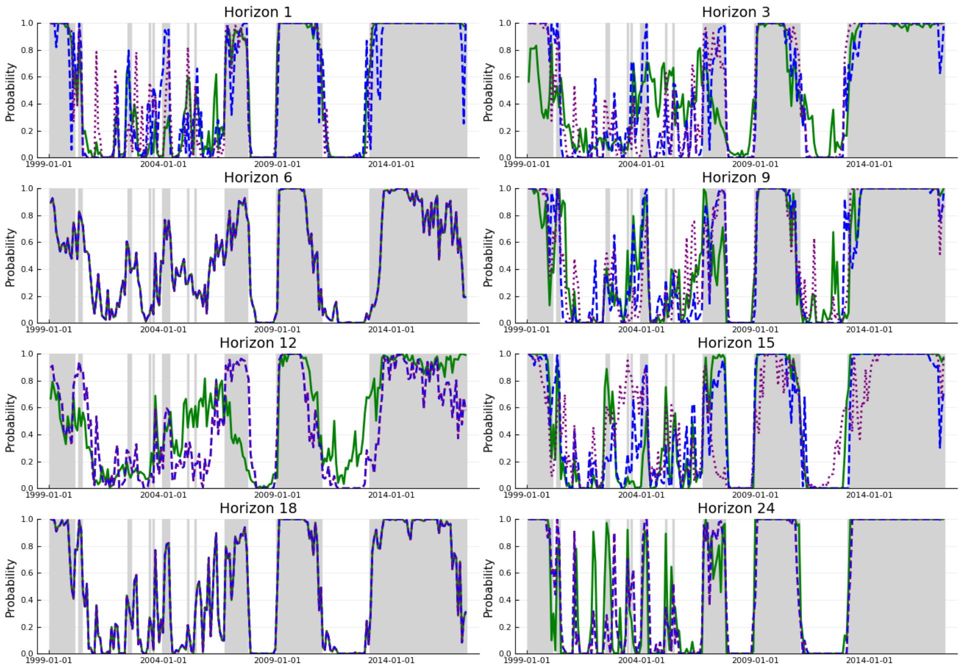

Figure 4: In-sample-model-fit for all horizons. Each panel shows the best model selected using AUROC (green solid lines), MAE (blue dashed lines) and RMSE (purple dotted lines). The grey bars represent periods with inflation smaller than 2%.

The models overall have a satisfactory in-sample prediction ability. Also, the models chosen with the three criteria are very similar, and on many occasions, the estimated probabilities overlap as for $h = 6$ and $h = 18$.

4.2 Out-of-sample analysis

After assessing the in-sample performance of our set of models, I evaluate their out-of-sample performance. The exercise performed is very similar to the one I used in in the first and second steps. For each set of variables, I pre-estimate the model from January 1999 to March 2007 and compute a direct forecast up to the end of the sample. Figure 5 shows the out-of-sample estimates of the different models for all the horizons. The sample-period I am using for the estimation presents two extended periods of inflation below the 2% level divided by a time of inflation above or equal 2%. These periods partially overlaps with the great recession (inflation below the target starts in December 2008) and to the post-European debt crisis (Jannuary 2013).

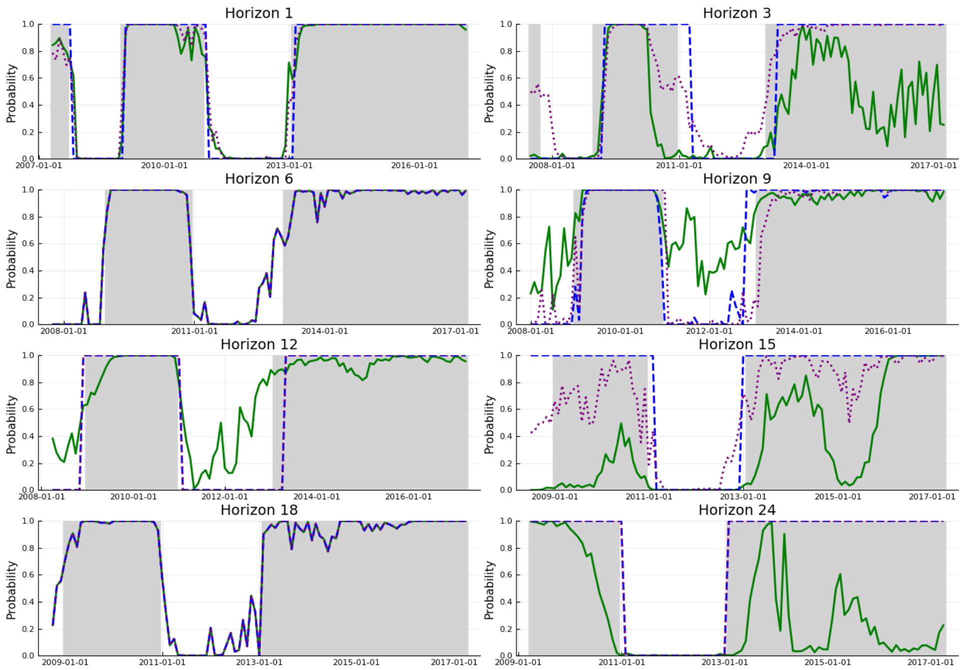

Figure 5: Out-of-sample forecast from probit mdel using AUROC (solid-green), MAE (dash-blue) and RMSE (dot-purple).

From visual inspection, the out-of-sample predictions do not show severe lack, and the overall fit is pretty good. However, as usual in out-of-sample forecasting, the goodness of the projection is a function of the horizons, and overall results are heterogeneous. At one-step-ahead, the three models have a performance which is exceptionally close. The models do an excellent job in timely catch the turning points in the inflation probability. At $h = 3$, the RMSE is the best model, succeeding in catching the first-period of low inflation and start rising slightly in advance with respect to the third one. The MAE model is somewhat delayed to the turning points, while the AUROC model delivers an extremely meager job. At horizon six and eighteen the three criteria have selected the same variables among all the possible combinations. This choice implies that the predictor selection is exceptionally robust across the different loss functions. For what concern the remaining horizons, the MAE and RMSE models are the best performers, and on many occasions, the variables selected by the two coincide. From the other side, the AUROC is not delivering a satisfying outcome, especially at horizons $h = 15$ and $h = 24$. The main problem seems to be the high autocorrelation in the estimated probabilities which miss the turning points. At the opposite, the other models do a solid job in timely capturing the switch from one state to the other.

Table 7: Results from out-of-sample forecast. All models, criteria and horizons.

	AUROC			MAE			RMSE		
Horizon	**AUROC**	**MAE**	**RMSE**	**AUROC**	**MAE**	**RMSE**	**AUROC**	**MAE**	**RMSE**
$h = 1$	1.02	0.99	1.02	0.46	0.25	0.41	0.68	0.73	0.64
$h = 3$	1.05	0.96	1.04	1.41	0.48	0.59	1.43	0.99	0.75
$h = 6$	1.19	1.19	1.19	0.17	0.17	0.17	0.35	0.35	0.35
$h = 9$	1.63	1.59	1.6	0.39	0.11	0.11	0.58	0.38	0.31
$h = 12$	1.62	1.57	1.57	0.3	0.09	0.09	0.46	0.37	0.37
$h = 15$	1.81	1.55	1.79	0.69	0.15	0.28	0.9	0.46	0.39
$h = 18$	1.54	1.54	1.54	0.12	0.12	0.12	0.21	0.21	0.21
$h = 24$	1.51	1.48	1.48	0.86	0.02	0.02	1.0	0.16	0.16

Note: the table shows the scores for each model selected according to AUROC, MAE, and RMSE and for each horizon h. The total number of selected model is 24, and each selected model is evaluated according to all the three possible criteria. The reported number is the ratio of the score of the selected model over the score of a univariate probit model which uses as regressor the first lag of the HICP inflation (naive model). For the AUROC, a score higher than one implies that the selected model outperforms the naive model, while for MAE and RMSE the opposite is true.

Table 7 summarizes the results of all the models, criteria, and horizons with respect to the naive model. The three panels show the results for the three criteria, while probit AUROC, MAE, and RMSE highlights the set of models chosen to maximize/minimizing these criteria. Therefore, by definition, for the AUROC panel, the set of models which attains the best results is the AUROC column, for the MAE panels is the MAE column and so on. It is interesting to notice that while MAE and RMSE models are very close to the AUROC models in terms of the score, the opposite is not true, and AUROC chosen models are very distant from MAE and RMSE models. Oddly, at $h = 3$, according to both the MAE and RMSE criteria the AUROC model is outperformed even from the naive model. The same happens to the very short horizons for the MAE in terms of AUROC. However, the distance between the two models is much smaller (one to three percentage points).

As a final exercise, I use all the estimated model to predict the probability of having low inflation in a true out of sample forecast. I predict all the horizons of interest starting from the end of the sample. I use each model to forecast only the single horizon for which the model is tailored. Table 8 shows the predicted probability of having low inflation in the 24 months after the model is calibrated. All the model predictions are very close, and clearly, a signal that the inflation upside risk does not

Table 8: Prediction from the best model at different horizons.

Date	AUROC	MAE	RMSE
Apr 2017	0.87	0.99	0.99
Jun 2017	0.99	0.95	0.95
Sep 2017	0.98	0.98	0.98
Dec 2017	0.99	0.98	0.97
Mar 2018	0.99	0.99	0.99
Jun 2018	0.99	0.99	0.99
Sep 2018	0.99	0.99	0.99
Mar 2019	0.99	0.99	0.99

seems a concern.

4.3 Deflationary pressure index

As a final exercise, in this section, I create an index to signal the probability of having low inflation in the medium run. I call the index *Deflationary Pressure Index* (DPI), given this is an index that signals the average probability of having inflation under the 2% level in the next two years. Due to the generic medium-term horizon considered by the European Central Bank to undertake policy actions, it could be useful to have a more general idea of the average inflation direction, in addition, to have many reference points at different horizons from many different models. Equation 12 shows the DPI. This is a simple average over the forecasts $h = [1, 3, 6, 9, 12, 15, 18, 24]$ coming from the AUROC, MAE and RMSE models.

$$DPI_t = \frac{1}{H}\sum_h P(\Pi_{T+h|T} = k|X_T, \pi_{t-1}), \quad k = L, H \tag{12}$$

Figure 6 shows the constructed deflationary pressure index against the periods in which inflation is below the 2% level. Ellipses shows the true out-of-sample forecast from the end of the sample to March 2019. The chart presents two obvious features; first, the movements in the three indexes are incredibly close. This characteristic is an excellent signal given that the three were created using three different loss functions and enhance the confidence in our forecast. Secondly, the indexes, accordingly to the true out-of-sample estimates, do not signal any upside risk for inflation in the next two years.

The central policy message covered by the index is that the adoption of a restrictive monetary policy measure to stabilize inflation in the next two years is very unlikely. In turn, this suggests that, as the economic conditions do not deteriorate, it would be very hard to see a hike in the interest rate in

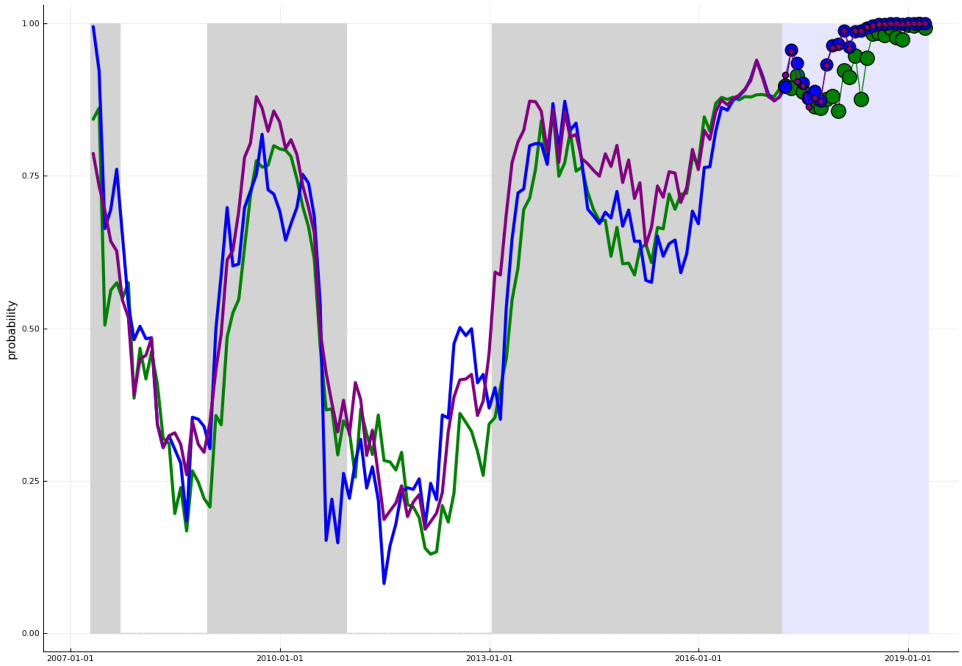

Figure 6: Deflation probability index from averageing the out-of-sample forecast for all horizons best models. AUROC (solid-green), MAE (solid-blue) and RMSE (solid-purple).

the following quarters. In turn, this implies that the era of negative rates will still last for a prolonged period. Of course, this does not suggest that the ECB can not move away from its historical rule and surprise the markets with a hike for reasons which are different from the inflation level. An example could be the willingness to restore a more standard monetary policy condition as soon as a stable inflation path is reached.

4.4 Survey of Professional Forecasters: a comparison

In this section, I compare the output of our model using as a benchmark the ECB *Survey of Professional Forecasters*. Starting from December 2000 the SPF are regularly collected from surveying more than 80 professional forecasters, who are the member of financial and non-financial institutions within the European Union. In a typical survey, they are required to express their point forecast about what they expect inflation (as well as GDP growth and unemployment) to be over specific time horizons. Also, they are asked to provide probabilities for different inflation outcomes. For example, they are asked to give the likelihood that the year-on-year HICP inflation is below, in between or above certain thresholds. The thresholds range from -1% to 4% stepping by 0.4%, for a total of 12 bins. Probabilities have to sum to one, and as the final forecast measure, the average of all the forecast among forecaster is used.

To create a SPF measure comparable to the deflationary pressure index, I construct a similar device by cumulating the probabilities of having inflation below the 2% level between -1% and 2% in the next 24 months. Also, given that the SPF are quarterly collected, I plot the indexes at mixed frequencies. Figure 7 shows the quarterly SPF median survey forecast (red-circle) against the monthly deflationary pressure index. I report The light blue zone highlights the data points for the true forecast. The chart shows that the SPF present more autocorrelation with respect to the DPI,

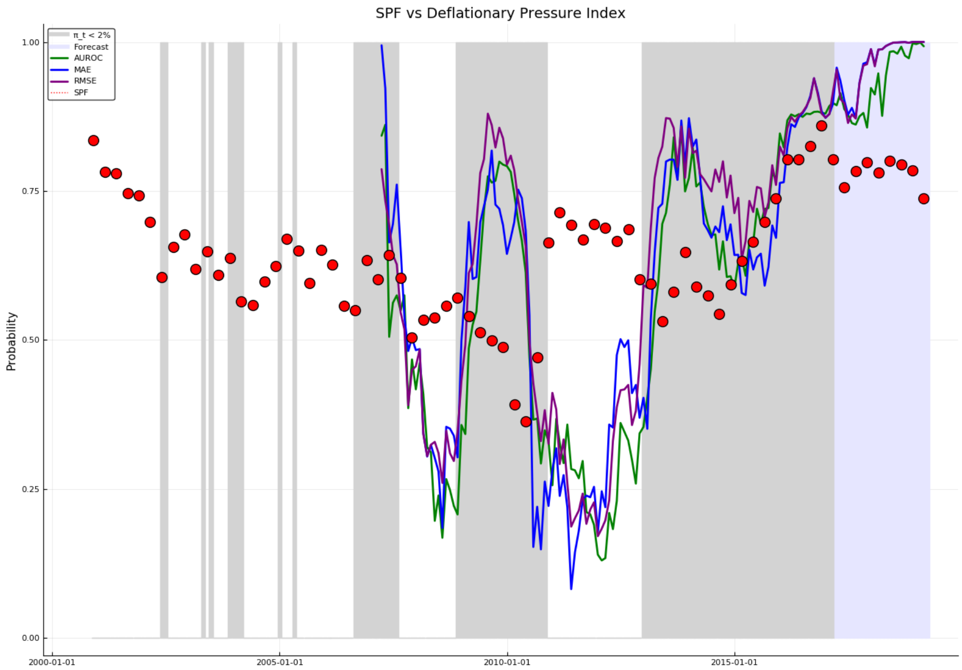

Figure 7: Deflation probability index from averageing the out-of-sample forecast for all horizons best models against SPF 24 months forecast. AUROC (solid-green), MAE (solid-blue), RMSE (solid-purple) and SPF (red dots).

being much slower in catching the turning points. The autocorrelation is especially evident in the last two transition period of inflation above/below the 2% threshold level. A part of that, the two measure looks close, especially starting from 2013. In the true out-of-sample forecast, both models predict a very high probability of having inflation below the 2% level. However, although they present a decline in probabilities in the second quarter of 2017, then the probability of the DPI increase again while the SPF prediction, after a slow increase starts declining at 75% level.

5 Conclusion

In this paper, I use a broad set of macroeconomic variables to forecast the probability of having inflation below the 2% level. To deal with the size of the dataset, I develop a two-step approach based on the combinatoric calculus empowered using the Julia language and exploiting a parallel computational approach on an octa-core processor. The main aim of the paper was to create an index to forecast the probability of having inflation below the 2% level over the next two year. To accomplish this task, I first select twenty-four different models using three different loss function (the AUROC, MAE, RMSE) and specialize each of them to forecast a particular horizon. Then I average the forecast from all the models to get a meaningful out-of-sample index. The index gives the probability of having inflation below the 2% level in the medium-run. The main idea is that an index capturing the probability of having inflation below a certain threshold can help in taking monetary policy decisions. In fact, central banks can be interested in the medium-run probability of deviating from the target as an additional measure to build confidence in their policy decision. In the present context, the index shows that in the next two years inflation is very likely to stay below 2%.

References

Berge, T. J. and Jordá, Ó. (2011). Evaluating the Classification of Economic Activity into Recessions and Expansions. *American Economic Journal: Macroeconomics*, 3(2):246–277.

Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM Review*, 59(1):65–98.

Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2009). New Keynesian Models: Not Yet Useful for Policy Analysis. *American Economic Journal: Macroeconomics*, 1(1):242–266.

Christoffel, K. P., Coenen, G., and Warne, A. (2008). The new area-wide model of the euro area: a micro-founded open-economy model for forecasting and policy analysis. *ECB Working Paper, 944*.

Dieppe, A., Pandiella, A. G., Hall, S. G., and Willman, A. (2011). The ECB's New Multi-Country Model for the euro area: NMCM-with boundedly rational learning expectations.

Dieppe, A., Pandiella, A. G., and Willman, A. (2012). The ECBs New Multi-Country Model for the euro area: NMCM Simulated with rational expectations. *Economic Modelling*, 29(6):2597–2614.

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

Estrella, A. and Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, 46(2):555–576.

Estrella, A. and Mishkin, F. S. (1998). Predicting U.S. Recessions: Financial Variables as Leading Indicators. *Review of Economics and Statistics*, 80(1):45–61.

Faust, J. and Wright, J. H. (2013). Forecasting Inflation. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2A, chapter 1, pages 2–56. Elsevier.

Galí, J. (2015). *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press.

Giannone, D., Lenza, M., Momferatou, D., and Onorante, L. (2014). Short-term inflation projections: A Bayesian vector autoregressive approach. *International Journal of Forecasting*, 30(3):635–644.

Liu, W. and Moench, E. (2016). What predicts US recessions? *International Journal of Forecasting*, 32(4):1138–1150.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1):19–46.

Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Stock, J. and Watson, M. (2016). Dynamic Factor Models Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. In Taylor, J. B. and Uhlig, H., editors, *Handbook of Macroeconomics*, volume 2A, chapter 8, pages 415–525. Elsevier.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press, Cambridge (Massachusetts).

Wu, J. C. (2017). Time-Varying Lower Bound of Interest Rates in Europe. *Chicago Booth Research Paper, No. 17-06*.

Wu, J. C. and Xia, F. D. (2016). Measuring the Macroeconomic Impact of Monetary Policy at the Zero Lower Bound. *Journal of Money Credit and Banking*, 48(2-3):253–291.

6 Appendix

6.1 Table

Variable	RIC-DS ID Code	Transformation Code
HICP EA	aXZCPIHICP/C	3
IP EA	aXZCINDG/CA	3
Consumers good	aXZPDAGCGS/A	3
Durable	aXZPDAGCDRB/A	3
Non durable	aXZPDAGCNDR/A	3
Intermediate	aXZPDAGINTG/A	3
Energy	aXZPDAGENE/A	3
Capital	aXZPDAGCAPG/A	3
Construction	aXZIPCON/A	3
Manufacturing	aXZIPMAN/A	3
Unemploy. rate EA	aXZUNR/A	1
Credit gen gov	aXZCRDGOV/A	3
Car regist	aXZCRDRG/A	3
Business climate	aXZBUSCLIM	6
Consumer confidence	aXZECOSE	6
Industrial confidence	aXZBSMFGCI/A	6
Retail confidence	aXZBSSVRTCI/A	6
Construction confidence	aXZBSCSCI/A	6
Service confidence	aXZBUCFM/A	6
Core CPI ea	aXZCCORF/C	3
Eonia	aXZONIA	1
M1	aXZM1	3
M2	aXZM2	3
M3	aXZM3	3
Neer	aXZINECE/C	3
US ff rate	aUSFEDFUND	1
IP DE	aDECINDG/A	3
IP ES	aESCINDG/A	3
IP FR	aFRCINDG/A	3

IP IT	aITCINDG/A	3
Unemploy. DE	aDECUNPQ/A	1
Unemploy. ES	aESCUNPQ/A	1
Unemploy. FR	aFRCUNPQ/A	1
HICP DE	aITUNRM/A	3
HICP ES	aESHICP	3
HICP FR	aFRHICP	3
HICP IT	aITHICP	3
Core CPI DE	aDECCORF/C	3
Core CPI FR	aESCCORF/C	3
Core CPI IT	aITCCORF/C	3
EA stock	.STOXX50E	3
EA bank stock	.SX7P	3
US stock	.SPX	3
US vol	.VIX	3
Crude	LCOc1	3
US10Y	US10YT=RR	1
EURIBOR3M	EURIBOR3MD=	1
EURIBOR6M	EURIBOR6MD=	1
EURIBOR1Y	EURIBOR1YD=	1
DE stock	.GDAXI	3
ES stock	.IBEX	3
FR stock	.FCHI	3
IT stock	.FTMIB	3
DE2YT	DE2YT=RR	1
DE5YT	DE5YT=RR	1
DE10YT	DE10YT=RR	1
ES5YT	ES2YT=RR	1
ES10YT	ES10YT=RR	1
FR2YT	FR2YT=RR	1
FR5YT	FR5YT=RR	1
FR10YT	FR10YT=RR	1
IT2YT	IT2YT=RR	1
IT5YT	IT5YT=RR	1

IT10YT	IT10YT=RR	1
NL2YT	IE2YT=RR	1
NL5YT	NL5YT=RR	1
NL10YT	NL10YT=RR	1
EA short repo	RC2AALM	1
EA2Y	EMECB2Y.	1
EA3Y	EMECB3Y.	1
EA5Y	EMECB5Y.	1
EA7Y	EMECB7Y.	1
EA10Y	EMGBOND.	1
Loans to nonfin corps	EMEBMC..A	3
Loans to hsld	EMEBMH..A	3
Loans to non-mfi	EMEBMEO.A	3
Loans to mfi	EMECBXLMA	3
IT CPI SA	ITCCPI..E	3
IT core CPI SA	ITCCOR..E	3
ES CPI SA	ESCCOR..E	3
DE CPI SA	BDCONPRCE	3
DE CPI core SA	BDUSFG10E	3
FR CPI SA	FRCONPRCE	3
FR core SA	FRCPUNDEE	3
Price trends 12M	EMZEWCP.R	6
Econ 12M	EKTOT4BSQ	6
Unemployment 12M SA	EKTOT7BSQ	6
REER	EMI..RECE	3
US crude	USSCOP.BP	3
Shadow	-	1