Working Paper Series —— 2/2025

Forecasting Disaggregated Producer Prices: A Fusion of Machine Learning and Econometric Techniques

Soňa Benecká

Czech National Bank —— Working Paper Series —— 2/2025

CZECH NATIONAL BANK

www.cnb.cz

Reviewed by: Joan Paredes (European Central Bank)

Ondřej Michálek (Czech National Bank)

Project Coordinator: Michal Franta

Forecasting Disaggregated Producer Prices: A Fusion of Machine Learning and Econometric Techniques

Soňa Benecká *

Abstract

This paper proposes a novel framework to the forecast of disaggregated producer prices using both machine learning techniques and traditional econometric models. Due to the complexity and diversity of pricing dynamics within the euro area, no single model consistently outperforms others across all sectors. This highlights the necessity for a tailored approach that leverages the strengths of various forecasting methods to effectively capture the unique characteristics of each sector. Our forecasting exercise has highlighted diverse pricing strategies linked to commodity prices, autoregressive behavior, or a mixture of both, with pipeline pressures being especially pertinent to final goods. Employing a mixture of a wide range of models has proven to be a successful strategy in managing the varied pricing behavior at the sectoral level. Notably, tree-based methods, like Random Forest or XGBoost, have shown significant efficacy in forecasting short-term PPI inflation across a number of sectors, especially when accounting for pipeline pressures. Moreover, newly proposed Hybrid ARMAX models proved to be a suitable alternative for sectors tighly linked to commodity prices.

JEL Codes: C22, C52, C53, E17, E31, E37.
Keywords: Disaggregated producer prices, forecasting, inflation, machine learning.

---

*Soňa Benecká, Czech National Bank, Monterary Policy Division, sona.benecka@cnb.cz
The views expressed in this paper are those of the authors and not necessarily those of the Czech National Bank.
The authors would like to thank Ondeřej Michálek and Joan Paredes for for their comments. Any errors and omissions are the responsibility of the authors.

1. Introduction

The Czech National Bank's primary modeling framework, the g3+ model (Brázdik et al., 2020), incorporates external price pressures through the euro area Producer Price Index (PPI). In the latest version of this model, producer prices are divided into core and energy components to enhance accuracy in economic forecasting and intepretation. The core component primarily reflects fundamental macroeconomic developments abroad, while the energy component is specifically designed to capture fluctuations in foreign production sectors that are closely linked to oil and other energy prices. Although variations in energy prices do not directly influence the price competitiveness of Czech exporters—most of whom compete in international markets for non-energy goods and services—both energy and non-energy price shocks have substantial impact on the Czech economy.

Properly distinguishing between energy and core prices in the model is critically important. Previously, the g3+ model primarily accounted for the direct impact of highly volatile energy prices, such as oil. However, following the Russian invasion of Ukraine and the subsequent threats of a sudden stop in gas deliveries, which propelled gas prices to unprecedented levels, a revision was necessary. Since 2022, the energy component of the euro area effective PPI has been expanded to include sectors related to oil, gas, and electricity. This revision was supported by an earlier version of the approach discussed in this paper, which aided the transition to a new framework by providing a detailed insights on futures developments in producer prices at a disaggregated level.

Forecasting producer prices at a disaggregated level presents considerable challenges and is an area that remains relatively underexplored in empirical literature. This paper aims to address several gaps in this field. Firstly, the model design incorporates a high degree of heterogeneity in price dynamics across individual sectors. This nuanced segmentation facilitates a more detailed analysis of producer price dynamics, enhancing our understanding of historical trends and future projections. Employing a *bottom-up* approach, the model can capture sector-specific shocks, thereby enriching the discussion on the impact of foreign price pressures.

Secondly, this study integrates the latest advances in machine learning—a methodology so far missing in the modeling and forecasting of producer prices. The goal here is not merely to determine whether machine learning methods can outperform traditional autoregressive models, but to leverage their distinct advantages, such as sophisticated variable selection procedures, in the quest for an optimal model. Our analysis reveals that producer price subgroups exhibit diverse price kernels: while some sectors are heavily influenced by commodity prices (both energy and non-energy), others adhere to a steady repricing pattern or respond to price pressures with a certain lag as these pressures propagate through the production chain. The optimal model setup, therefore, should recognize such variances, selecting appropriate explanatory variables or formulating an independent price kernel accordingly.

Moreover, as a novel contribution, we conduct a comprehensive evaluation of a broad array of machine learning methods to assess their forecasting performance. We introduce several innovative methods that synergize the strengths of traditional econometric approaches with the capabilities of modern machine learning techniques. These hybrid models are designed to leverage the robust analytical frameworks of econometrics while incorporating the adaptive learning and predictive power of machine learning, thus enhancing their effectiveness in forecasting complex economic phenomena.

Moreover, this study explores how variations in the level of data aggregation affect our forecasting results. A bottom-up approach, while detailed, may not always be adequate due to substantial noise inherent in the data. We investigate whether utilizing different levels of granularity can improve the accuracy and reliability of our forecasts by mitigating the noise associated with broader category analyses. Additionally, we aim to provide empirical evidence on the significance of "Pipeline Pressures" from a forecasting perspective. This term describes a cascade of sectoral price shocks that propagate through the production chain, impacting subsequent stages and ultimately influencing overall economic outcomes. Understanding these pressures is crucial for developing more accurate predictive models that reflect the interconnected nature of production sector.

Finally, this research provides a robust framework that could be utilized in forecasting exercises by the Czech National Bank. Although the Czech National Bank traditionally considers foreign variables such as GDP or PPI in effective terms, a complementary assessment of price pressures in the euro area, as detailed in this study, can offer valuable insights.[1]

This document is structured into five main sections. The second section offers a concise review of the literature pertinent to this paper, setting the stage for the subsequent analysis. The third section provides an overview of the various models employed throughout the study, detailing their theoretical foundations and operational mechanisms. The fourth chapter presents data and the empirical findings from our forecasting exercises. It identifies the most effective forecasting methods and discusses their insights into the principal drivers of the producer prices in the euro area. A subchapter explores the advantages of utilizing forecast combinations and model selection at a disaggregated level. Additionally, a special section examines whether a lower level of aggregation impacts the accuracy of our results. The final robustness check assesses the influence of pipeline pressures on forecasting disaggregated producer prices, providing evidence of their significance in the predictive process. The fifth and final chapter summarizes the conclusions and lessons learned from this research.

[1] The Producer Price Index in effective terms is computed as a weighted average of PPIs from euro area countries. The weighting scheme is based on the share of each country in Czech exports.Given the high degree of synchronicity in PPIs across European countries, the differences between this aggregate index and one based solely on the euro area are relatively minor. The model's design is also adaptable enough to be applied to data from individual countries. However, Eurostat's confidentiality restrictions prevent its application at the detailed NACE Rev. 2 level. Nonetheless, the model can be readily adapted for use at the Main Industrial Grouping (MIG) disaggregation level, where such data restrictions do not generally apply.

2. Literature

Machine learning (ML) methods have gained substantial traction in the field of economics, particularly for inflation forecasting. Recent studies demonstrate that ML methods surpass traditional econometric models in forecasting accuracy and efficiency. This superiority is attributed to the ML methods' capability to discern nonlinear relationships and intricate patterns within the data, which are often overlooked by conventional models. Additionally, ML techniques are particularly advantageous in scenarios involving high dimensionality, where the number of predictors exceeds the number of observations, or where interactions among variables are complex. A pivotal strength of ML methods is their empirical approach to managing the bias-variance trade-off, for example through regularization, a process thoroughly discussed in the work of Kleinberg et al. (2015). Regularization helps in optimizing prediction performance by effectively balancing model complexity and training data fit, thereby enhancing the model's generalization capabilities in unseen data.

However, the deployment of machine learning methods in inflation forecasting is fraught with challenges. A principal concern is the interpretability of results, as these methods are frequently perceived as "black boxes." This opacity complicates the understanding and validation of the models' decision-making processes. Additionally, the quality of forecasts is susceptible to variations in data quality, the selection of features, and the specific configurations of the models. Despite these obstacles, the body of literature on the application of ML methods to inflation forecasting is rapidly evolving. Current research is primarily focused on enhancing both the accuracy and the explainability of these models, aiming to bridge the gap between technical performance and practical usability.

A variety of machine learning techniques have been employed to forecast consumer inflation, including penalization regressions, random forests, gradient boosting, and artificial neural networks. Horse race studies generally favor random forests in several analyses. Chakraborty and Joseph (2017) assessed the effectiveness of ten econometric and machine learning models for predicting inflation in the United Kingdom, noting a significant decline in predictive accuracy following the Global Financial Crisis (GFC). Before the GFC, support vector machines (SVM) excelled in terms of mean absolute error over a two-year forecasting horizon. However, post-GFC, random forests emerged as the most effective stand-alone model, although a hybrid model combining SVM with a deep neural network surpassed all individual approaches in performance. Medeiros et al. (2021) also advocates for random forests as a robust forecasting tool for U.S. inflation, attributing its superior performance to its variable selection technique and its capacity to capture nonlinear relationships between key macroeconomic factors and inflation. This study further suggests that the application of machine learning techniques in inflation forecasting could enhance prediction accuracy by as much as 30% in terms of mean squared errors. Kohlscheen (2021) utilized regression trees to predict inflation across twenty advanced countries from 2000 to 2021, offering insights into the underlying drivers of inflation.Araujo and Gaglianone (2022) conducted a comprehensive analysis encompassing a broad range of models and variables, concluding that the most accurate forecasts typically arise from combinations of models, tree-based methods such as Random Forest and XGBoost, breakeven inflation rates, and survey-based expectations. Conversely, Garcia et al. (2017) found that LASSO-based models outperformed random forests, and that combining models could yield superior predictive results. More recently, Schnorrenberger and Moura (2024) demonstrated that the LASSO method for selecting the best predictors is optimal for nowcasting Brazilian inflation.

Moreover, there is increasing evidence supporting the efficacy of simple neural networks in forecasting inflation, suggesting their potential as a tool in economic prediction, for example Nakamura (2005); Moshiri and Cameron (2000); Choudhary and Haider (2012). Also deep learning methods have garnered significant attention in the area of macroeconomic forecasting. Notably, the integration of Variational Autoencoders with Convolutional LSTM Networks (VAE-ConvLSTM) has proven particularly effective, surpassing several established econometric and machine learning benchmarks, including random forests (Theoharidis et al., 2023). Paranhos (2021) highlights the advantages of using LSTM for long-term forecasting horizons, underscoring its capability to capture long-term inflation trends effectively. Almosova and Andresen (2023) demonstrates that while Long Short-Term Memory networks marginally outperform autoregressive models, other neural networks, and Markov-switching models, they achieve comparable performance to the Seasonal Autoregressive Integrated Moving Average model.

To date, only a limited number of studies have examined disaggregated consumer price data, which illuminates the bottom-up approach to forecasting inflation. Boaretto and Medeiros (2023), in preliminary version, validates the effectiveness of machine learning methods in forecasting subgroups of the Brazilian inflation, demonstrating that improvements in these subgroup forecasts enhance the accuracy of the overall index prediction. Particularly in the post-COVID period, the study highlights the robust predictive performance of random forests over medium to long forecasting horizons. Similarly, Barkan et al. (2022) employs recurrent neural networks to simultaneously forecast components of the U.S. consumer prices. This method leverages information from higher levels in the hierarchy to refine predictions at the more volatile lower levels, showcasing the potential of hierarchical data integration in enhancing forecasting precision.

While considerable attention has been paid to consumer inflation, producer inflation has been somewhat underexplored in the economic literature.[2] One primary focus within this lesser-studied area is the investigation of cross-country linkages in production and price spillovers. Auer et al. (2019) demonstrates that cross-border propagation of cost shocks, facilitated through input-output linkages, plays a significant role in synchronizing producer price inflation across countries. More recently, LaBelle and Santacreu (2022) explored how foreign bottlenecks influence the transmission of supply chain disruptions to U.S. producer prices in the aftermath of the COVID-19 pandemic. The shift in consumer demand towards durable goods, combined with a heavy reliance on foreign suppliers for these products, has led to a mismatch between supply and demand, resulting in a surge in prices. Sectors that depend extensively on foreign inputs from countries experiencing greater disruptions have seen more pronounced increases in producer inflation.

A strand of research has been focused on the dynamics and underlying drivers of producer inflation, utilizing firm-level data to uncover nuanced insights. A study by Gautier (2008) investigates price-setting behaviors among French companies, revealing that price adjustments occur quite frequently and exhibit no asymmetry between increases and decreases. The study uncovers significant sectoral heterogeneity: while business-service prices change less often, the adjustments are generally larger compared to those in the industrial sector. Firms typically modify their prices annually or biennially, predominantly in January, aligning with Taylor's theory of staggered price contracts. The analysis also indicates that price changes are influenced by sectoral inflation, phases of the business cycle, and raw material costs. Research on price stickiness in Belgian producer prices by Cornille and Dossche (2006) demonstrates that producer prices undergo more frequent yet smaller adjustments, suggesting relatively low price adjustment costs for producers. The study reveals a significant variation in the frequency of price adjustments across different sectors, whereas the magnitude of these

[2] The connection between consumer prices and producer prices has garnered focused attention in economic research. A limited number of studies have explored the causal dynamics between these two price indices. For instance, Sidaoui et al. (2010) and Wei and Xie (2018) have examined the predictive relationships between consumer and producer prices, providing insights into their interdependencies.

adjustments exhibits less variability. This finding implies that the frequency of adjustments, rather than their magnitude, primarily accommodates sectoral characteristics in pricing strategies. Interestingly, the study also indicates that differences in price adjustment costs do not significantly influence the observed heterogeneity in price stickiness. Instead, it is the variations in firms' cost structures and market dynamics—such as a high proportion of energy inputs, a lower share of labor costs, and intense market competition—that predominantly drive the frequency of price adjustments.[3]

The academic literature specifically addressing the forecasting of producer prices is sparse. Zhao et al. (2020) evaluated the efficacy of artificial neural networks (ANNs) against autoregressive integrated moving average (ARIMA) models for predicting the Producer Price Index in New Zealand. Their findings indicate that ANNs surpass ARIMA in terms of both reliability and accuracy for time series data prediction, offering a more effective tool for economists and macroeconomic policymakers. Additionally, McAdam and McNelis (2005) explores both linear models and neural network-based thick models for forecasting inflation. The concept of "thick" models involves generating trimmed mean forecasts aggregated from multiple neural network predictions, applied to Phillips-curve formulations in various economies, including the USA, Japan, and the euro area. The study demonstrates that these models are effective for forecasting both consumer and producer price indices across a broad set of countries, in both real-time and bootstrap forecasting scenarios. The combination of multiple model outputs into a single thick model provides a robust method for enhancing forecast accuracy in complex economic settings with fluctuating inflation dynamics. Furthermore, Bhattacharya and Thomakos (2008) implements VARX models incorporating exchange rate pass-through to augment the forecasting performance for industry-level producer inflation. Despite these contributions, there appears to be a gap in the literature concerning the comprehensive evaluation of forecasting methods for producer prices utilizing various machine learning techniques, compared to traditional forecasting tools.

3. Machine Learning and Econometric Models

Traditional forecasting methods, such as ARIMA and exponential smoothing, typically rely on establishing explicit relationships between dependent and independent variables, often requiring assumptions about specific functional forms and stochastic processes. These methods excel in contexts where the underlying process theories are well-understood but may fall short in capturing complex, nonlinear relationships present in the data. In contrast, machine learning (ML) offers a primarily data-driven approach, employing a diverse array of techniques—ranging from regression and decision trees to neural networks and advanced ensemble methods like random forests and gradient boosting machines. This approach minimizes assumptions about the data's statistical relationships, focusing instead on identifying patterns directly from the data.

The two key elements of ML are *the learning method*—the procedure through which a model iteratively adjusts its parameters to minimize prediction errors, exemplified by techniques such as gradient descent—and *the algorithm*—a specified set of rules or procedures that the model follows to learn from data, facilitating the automation of modeling decisions. This automation not only reduces the need for manual intervention by the forecaster but also adapts efficiently to various data characteristics, although it hinges significantly on the availability and quality of data. Furthermore, ML methods stand out when modeling multiple variables simultaneously, as they can adeptly select

[3] The role of producer prices in mediating commodity shocks has been highlighted by for example Myers et al. (2018), that reveals significant long-run connections among oil prices, producer prices, and consumer prices in response to permanent shocks. Furthermore, the impact of producer prices on exchange rate pass-through has been studied by Jiang and Kim (2013), underscoring their importance in international trade dynamics.

and tune different models for each variable, capturing complex interdependencies that traditional methods have difficulties to capture.[4]

- **Supervised Learning** is the most prevalent form, where the goal is to map input data to known outputs, refining the model through comparison of its predictions against actual outcomes. Techniques such as linear regression, including penalization, are used for continuous outcomes, while logistic regression is applied for categorical outcomes. Ensemble methods (Random Forest, Gradient Boosting) create a group (or "ensemble") of models through sampling or improvement ("boosting"). Finally, k-Nearest Neighbors (k-NN) algorithm predicts based on the labels or values of the neighboring points.
- **Unsupervised Learning** deals with data without labeled responses or continuous outcome variable. The aim here is to shed a light on underlying structures or distributions in the data, assigning it into clusters or reducing its dimensionality for better analysis. Clustering analysis, like K-means, and dimensionality reduction techniques, such as principal component analysis, are common examples.
- **Semi-Supervised Learning** bridges the gap between the supervised and unsupervised approaches. It is particularly useful when acquiring a fully labeled dataset is too expensive or time-consuming, but unlabeled data is plentiful. By leveraging a small amount of labeled data alongside a larger set of unlabeled data, semi-supervised learning can improve learning accuracy. Techniques often involve a combination of supervised methods to label the unlabeled data, which then informs further learning, as seen in methods like self-training and co-training.
- **Reinforcement Learning** is distinct from the other types, focusing on making a sequence of decisions. The learning system, called an agent, learns to achieve a goal in an uncertain, potentially complex environment. It operates through trial and error, using feedback from its own actions and experiences, rather than being explicitly taught. Applications include robotics, gaming, and navigation, where the system optimizes for a cumulative reward. Q-learning, a popular algorithm in reinforcement learning, enables an agent to learn the value of actions in states of the world.

In this paper, we employ a selected set of *supervised* machine learning methods forecast producer prices. These methods include regularization techniques, which help in mitigating overfitting; regression trees, adept at modeling nonlinear relationships; neural networks, able to capture complex patterns; and support vector machines, which perform well in high-dimensional setup. This set represents a good starting point for constructing a robust forecasting framework centered around ML techniques. Following this, we complement these modern approaches with traditional econometric models, from autoregressive models and their extensions to factor models or Bayesian model averaging. While the array of potential ML models is expansive, our selection serves as a starting comparison to propose a comprehensive forecasting framework that integrates the strengths of both ML techniques and traditional approaches. At the end of this chapter, we also propose some innovations bridging traditional and new methods, exploting the advantages of both contemporaneously.

Before moving to the description of individual methods, let‘s summarise the key notation. Our aim is to forecast producer price inflation at time $t+h$ given a set of predictors at time t:

$$y_{t+h} = F_h(x_t) + \varepsilon_{t+h} \tag{1}$$

[4] The current literature on machine learning is abundant. For introduction see e.g. James et al. (2013).

The set of $n = 1...N$ predictors in x may contain weakly exogenous predictors, lagged inflation values and factors from the set of variables, if required. Alternatively, we consider also models, where expected predictors are available:

$$y_{t+h} = F_h(x_{t+h}) + \varepsilon_{t+h} \tag{2}$$

To evaluate forecast, we split the sample in two consecutive time sub-periods. The *estimation set* is the first sub-period $t = 1, ...T_1$ and it is used to estimate model in the recursive way, i.e. with increasing sample as we move along the forecasting exercise. The pseudo real-time out-of-sample forecasts are run on *evaluation set*, i.e. on the last sub-period $t = T_1 + 1, ...T$. Here we compute the root mean-squared error:

$$RMSE_h = \sqrt{\frac{1}{T - (T_1 + h) + 1} \sum_{t=T_1+h}^{T} \left(y_t - \hat{y}_{t|t-h}\right)^2}. \tag{3}$$

3.1 Benchmark and AR-based Models

One of the contributions of this paper is the comprehensive comparison - a 'horserace' - among forecasting methods, beginning with the fundamental benchmark of the **Random Walk (RW)** model. In this model, forecasts for a given horizon h are simply $y_{t+h} = y_t$, assuming future values follow the last observed value.

We further evaluate several AR-based models, which have shown particular strength in forecasting short-term trends within volatile time series. According to various studies (Almosova and Andresen, 2023), traditional time series methods with autoregressive components not only stand their ground against more contemporary approaches but also continue to represent a gold standard for forecasters.

Our first AR-based benchmark is the straightforward **Autoregressive Model (AR)** of order p, chosen based on the Bayesian Information Criterion (BIC). The model forecasts future values as $y_{t+h} = \phi_0 + \phi_1 y_t ... + \phi_p y_{t-p+1}$, incorporating the influence of past values up to p steps back.

To broaden our comparison, we examine the **Autoregressive Moving Average (ARMA)** model, which includes lagged forecast errors in its predictions. It can also automate the selection of autoregressive (p) and moving average (q) terms. The ARMA model's general formula is given by $y_t = \mu + \phi_1 y_{t-1} ... + \phi_p y_{t-p} + \theta_1 e_{t-1} ... + \theta_q e_{t-q}$. ARMA models are effective for forecasting inflation as they leverage the autoregressive part to capture the correlation between current and past rates, while the moving average component addresses short-term volatility and noise. ARMA model can be extended with additional explanatory variables to form **Autoregressive Moving Average with Exogeneous Variables (ARMAX)** model.

3.2 Penalized Regression Methods

The first set of simple methods fits generalized linear and similar models via penalized maximum likelihood. The most commonly used penalized regression methods include Ridge Regression, LASSO and Elastic Net. Generally, for linear models where $F_h(x_t) = \beta_h' x_t$ the shrinkage estimators can be defined as:

$$\hat{\beta}_h = \arg\min_{\beta_h} [\sum_{t=1}^{T-h} (y_{t+h} - \beta_h' x_t)^2 + \sum_{i=1}^{N} p(\beta_{h,i}; \lambda, \omega_i)] \tag{4}$$

where $p(\beta_{h,i}; \lambda, \omega_i)$ is a penalty function, depending on penalty parameter λ and on a weight ω_i, if required.

Ridge Regression (RR): Ridge Regression, with its origins in the 1970s as detailed by Hoerl (2020), uses coefficient shrinkage to diminish the influence of less significant variables, effectively bringing their coefficients nearer to zero, yet uniquely retaining all predictors within the final model. This method aims to mitigate overfitting by introducing a penalty term, the L2-norm, which is the sum of the squared coefficients. The extent of shrinkage is governed by the regularization parameter (λ), with its optimal value often determined through cross-validation. The penalty is hence given by:

$$\sum_{i=1}^{N} p(\beta_{h,i}; \lambda) = \lambda \sum_{i=1}^{N} \beta_{h,i}^2 \tag{5}$$

Lasso model (LASSO): Introduced by Tibshirani (1996), LASSO has emerged as a pivotal technique for models that benefit from shrinkage and variable selection. It reduces regression coefficients towards zero by imposing an L1-norm penalty, which is the sum of the absolute values of the coefficients, effectively allowing for the elimination of irrelevant variables by minimizing the residual sum of squares subject to the constraint that the sum of the absolute values of the coefficients is less than a certain constant. This mechanism not only facilitates model simplification but also aids in feature selection in high-dimensional datasets by promoting sparsity, where only a subset of predictors contributes to the model. The penalty imposed by LASSO is given by:

$$\sum_{i=1}^{N} p(\beta_{h,i}; \lambda) = \lambda \sum_{i=1}^{N} |\beta_{h,i}| \tag{6}$$

LASSO is particularly effective when the model includes predictors with large coefficients alongside others with negligible impact. In contrast, Ridge Regression is more suitable for scenarios where many predictors contribute more uniformly to the outcome (Gareth et al., 2013).

Elastic Net (ELNET): Introduced by Zou and Hastie (2005), Elastic Net combines the regularization features of both the L1-norm and L2-norm, offering a versatile approach to regression modeling. This method achieves coefficient shrinkage (akin to Ridge Regression) and variable selection (characteristic of LASSO) simultaneously. Its penalty is structured as:

$$\sum_{i=1}^{N} p(\beta_{h,i}; \lambda) = \alpha\lambda \sum_{i=1}^{N} \beta_{h,i}^2 + (1-\alpha)\lambda \sum_{i=1}^{N} |\beta_{h,i}| \tag{7}$$

Here, the tuning parameter α, which ranges from zero to one, determines the balance between L1 and L2 regularization, offering flexibility to adapt to various data characteristics. Elastic Net is particularly advantageous in situations with a large number of correlated predictors, as it can handle multicollinearity more effectively than LASSO or Ridge Regression alone. This makes it invaluable in scenarios where the predictors outnumber observations, a common challenge in high-dimensional data analysis. By leveraging the dual penalties, Elastic Net facilitates both automatic variable selection and continuous shrinkage, enabling a more nuanced model optimization process that can accommodate complex dataset structures.

Adaptive LASSO (ALASSO): Zou (2006) introduced an enhancement to the LASSO method by incorporating adaptive weights ω_i for penalizing coefficients differently. These weights, derived from initial estimates, allow for a tailored penalization, imposing less penalty on coefficients deemed more significant. This adaptive feature enables the method to maintain a balance between model sparsity and predictive accuracy, especially in situations with more predictors than observations and under conditions of heteroscedasticity Medeiros and Mendes (2016). The initial estimation step, based on OLS or standard LASSO, informs the adaptive weights, optimizing the selection of variables by penalizing the less relevant predictors more strongly:

$$\sum_{i=1}^{N} p(\beta_{h,i};\lambda,\omega_i) = \lambda \sum_{i=1}^{N} \omega_i |\beta_{h,i}| \tag{8}$$

3.3 Support Vector Regression

Aside penalized regression models, **Support Vector Regression (SVR)** relies on minimizing the **l2-norm** of the coefficient vector, similarly to the penalty in Ridge Regression. Moreover, it handles the error term through its constraints rather than minimizing the squared error directly. This methodology sets absolute errors less than or equal to a predefined margin (maximum error, ε). So it is a mixture of mean absolute error (MAE) regression with a ridge-like regularization approach. Such a dual focus not only minimizes the model's complexity but also tailors the error tolerance to specific requirements, distinguishing SVR from traditional regression techniques.

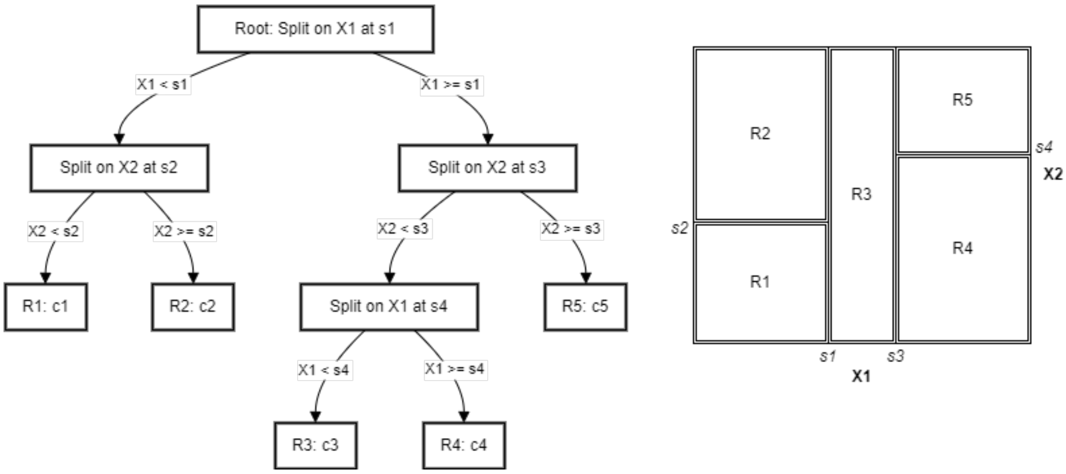
Support Vector machines are mostly used for classification, even if they offer an interesting alternative to regression, offering robustness against outliers by concentrating on errors exceeding the a certain threshold and thereby ensuring model stability. Working with absolute error it ignores outliers, so it relies on more stable model reflecting underlying data. Second, as the error margin helps us to ignore the data points that have small errors, it can speed up the model training, especially in computation heavy tasks. This feature, as discussed in the comprehensive technical background by Gunn et al. (1998), makes SVR an advantageous tool for large-scale or real-time data analysis. Still, its application in fields like inflation forecasting remains rare, with notable exceptions such as the study by Ye et al. (2013).

3.4 Random Forest

As highlighted in our literature review, **Random Forest Regression (RFR)** has emerged as a highly effective method for forecasting consumer inflation. This regression algorithm, an extension of the original random forest technique proposed by Breiman (2001) for classification, leverages an ensemble of decision trees to predict continuous numerical values. In RFR, the ensemble approach combines multiple decision trees, each constructed from randomly selected subsets of the training data and features. The decision-making process within each *regression tree* involves recursive binary splitting of the covariate space X, with the final prediction being the average of the outputs from all trees in the forest. This random selection process not only de-correlates the trees, enhancing model robustness, but also, through bootstrap aggregating (*bagging*), effectively lowers forecast variance without substantially increasing bias. Finally, parameter tuning such as adjusting the number of trees and/or their depth should be always considered, as it lowers computational burden and improves RFR performance.

To show the mechanism of a regression tree, we reference an example from Hastie et al. (2009), where X_1 and X_2 are explanatory variables and Y is a dependent variable. Initially, as depicted in Figure 1, the covariate space is split at $X = s_1$ into two regions, and this process is repeated, resulting in a total of five distinct regions ($R_k, k = 1,..5$), or terminal nodes. Each split is made to minimize the variance within each region, aiming at the model's predictive accuracy. In this simplified model, the value of Y in each region R_m is predicted with a constant c_m, calculated as the average of values Y for all observations within R_m, for $m = 1,..5$.

Figure 1: Example of Regression Tree Based on Hastie et al. (2009)



More formally, expected value of Y, the dependent variable, for any given values of explanatory variables X_1 and X_2 is a sum of constant values c_m assigned to each of the five terminal nodes and indicator function l that equals 1 if the observation falls within the region R_m, and 0 otherwise:

$$E_{regressiontree}(Y|(X_1,X_2)) = \sum_{m=1}^{5} c_m l_{(X_1,X_2)\in R_m} \tag{9}$$

3.5 Quantile Regression Forest

An important extension of Random Forest was developed by Meinshausen and Ridgeway (2006), **Quantile Regression Forest (QRF)**. It offers non-parametric and comprehensive estimates of conditional quantiles, including the median, for predictive analysis. By doing so, QRF enables more detailed assessment of the potential variability in outcomes, compared Random Forest, which focus on mean predictions. The methodology allows for the construction of prediction intervals that capture the dispersion of the data, thereby offering a robust evaluation of risks associated with extreme forecast outcomes. For example, QRF can be invaluable in financial risk management or climate modeling, where understanding the range of potential future scenarios is as critical as the prediction of central tendency.

The construction of this algorithm starts in a similar way as in Random Forest. Each tree partitions the data into sub-samples until a particular criterion, like mean squared error, is met or a minimum number of data points is reached, via binary recursive partitioning. After the final split of the data, the forecasted outcome for the target variable within a particular sub-sample is denoted by either the average or median value of that variable in the subgroup, which is referred to as a "leaf" when making point predictions. For density predictions, as opposed to point predictions, QRF evaluates empirical quantiles of the target variable associated with each leaf.

In order to compare predictive potential across diffent methods with point forecast, we must adopt alternative approach similar to Araujo and Gaglianone (2022). When constructing point forecast, we can still use information from conditional quantiles to achieve an improved conditional mean. Araujo and Gaglianone (2022) utilizes the mean of the quantile forecasts to construct their predictions to capture a more representative outcome across the quantiles. This method is particularly relevant in periods of economic uncertainty, such as during an energy crisis, where inflation rates were subject to unexpected spikes that often challenge forecasters. According to Lenza et al. (2023), QRF density forecasts as a new tool of ECB inflation forecasting toolkit proved better accuracy at the end of 2022, than than those from ECB Survey of Professional Forecasters.

More specifically, we follow this simplified algorithm:

1. Estimate Quantile Regression Forest on a set $i = 1...N$ of explantory variables $x_{i,t}$, given our response variable y_t.
2. For a fixed grid of quantile grid $R = [0.05, ...0.95]$ compute conditional quantile $Q_r(y_{t+h})$ for forecast horizon h.
3. Compute the mean across all conditional quantiles $Q_r(y_{t+h})$ and consider it as a forecast.

3.6 Gradient Boosted Regression Trees

Gradient Boosted Regression Trees model, pioneered by Friedman (2001) among others, belongs to the ensemble learning family as Random Forest, but there are some notable differences. It combines the predictions from multiple decision tree models to improve the overall predictive accuracy compared to what could be achieved by any single tree model. It uses the gradient descent algorithm to minimize the loss when adding new models. Without going into technical details, the algorithm iteratively add decision trees to predict and correct errors made by the ensemble so far. Each new tree focuses on the residual errors of the previous trees, guided by the gradient descent method to minimize the loss function. To prevent overfitting shrinkage or regularization are used by scaling down predictions and penalizing complexity. The process continues until a specified number of trees is reached or no further improvement is possible.

Gradient Boosting Trees and Random Forest differ primarily in their training approach and output determination. Gradient Boosting Trees are trained sequentially to correct prior errors, preventing parallel training, while Random Forest builds trees independently, allowing for parallel processing. Additionally, Random Forest aggregates independent tree outputs for decision-making, either by majority voting for classification or averaging for regression, whereas Gradient Boosting Trees rely on a fixed sequence of trees for sequential evaluation.

The diffence in outcomes between Gradient Boosting Trees and Random Forest can be particularly significant in case of correlated features (explanatory variables). In Random Forest, because each tree makes decisions independently, a perfectly correlated pair of features (A and B) will be split in importance roughly evenly across the trees, potentially obscuring the true value of this information. In contrast, Gradient Boosting Trees tend to concentrate importance on one of the correlated features, since the algorithm learns to avoid redundancy as it progresses, offering a clearer signal of feature importance but requiring further analysis to identify all relevant correlated features. On the other hand, both methods are suitable for problems where relationships between variables are nonlinear.

Using Gradient Boosted Regression Trees requires careful approach. Its performance can significantly depend on the tuning of hyperparameters such as the number of trees, depth of trees, learning rate, and regularization terms. Efficient hyperparameter optimization techniques, such as grid search, random search, or more advanced methods like Bayesian optimization, are crucial for maximizing model performance. Gradient Boosted Regression Trees are thus not immune to overfitting if not carefully regularized or if trained with too many trees. Therefore, a robustness check on hyperpameter tuning should be always included in any study deploying this method.

Several variations of Gradient Boosting Trees have been developed to improve performance, scalability, and usability. The most important implementations include XGBoost, LightGBM, and CatBoost. Each of these brings certain advantages, such as speed improvements, handling categorical features without preprocessing, and reducing overfitting. We will focus on two of them, namely XGBoost and CatBoost.

XGBoost (XGB) stands for eXtreme Gradient Boosting is well known implementation, as it includes also built-in routines for handling missing values, regularizing to prevent overfitting, and efficiently processing sparse data. It can handle large datasets and is optimized for speed and resource efficiency, top-ranking in speed among similar algorithms. **CatBoost (CTB)**, i.e. Gradient Boosting on Decision Trees, is especially powerful with cathegorical data and does not require preprocessing. It utilizes ordered boosting, a permutation-driven method that reduces overfitting and improves model robustness, but its performace in terms of speed is somewhat lower compared to XGBoost.

3.7 Neural Network

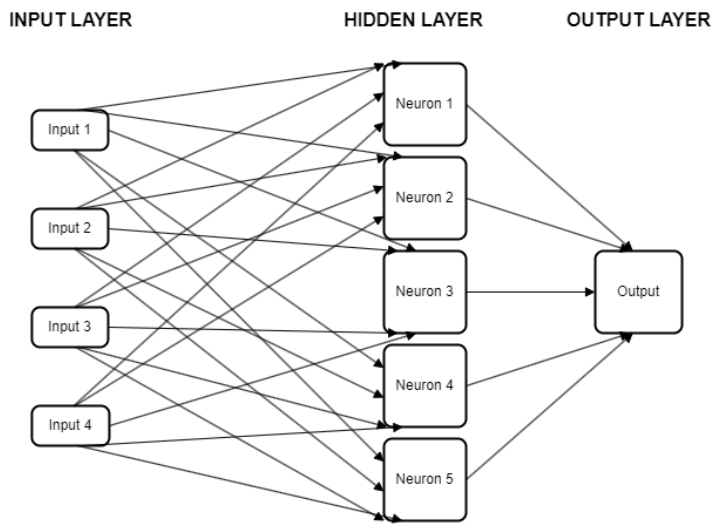
Several studies, like Almosova and Andresen (2023) or Barkan et al. (2022), suggest that artificial neural networks (ANN) can predict complex economic time series, including inflation. Artificial neural networks are forecasting techniques inspired by the brain's structure, utilizing simple mathematical models to manage complex nonlinear relationships between predictors and response variables.These networks are composed of various layers: the input layer represents predictors, hidden layers serve as computational engines that introduce nonlinearity to the system, and the output layer delivers forecasts.

The simplest ANN models lack hidden layers and linearly combine inputs using adjustable weights to generate predictions, as in linear regression. The optimization of these weights occurs through a learning process aimed at minimizing error metrics such as the mean squared error. The inclusion of one or more hidden layers transforms ANNs into nonlinear models capable of capturing the complex patterns in data. This transformation is exemplified in multilayer feed-forward networks (FNNs), where data flows unidirectionally from input to output, without any feedback loops, allowing each input to be processed independently.

In FNNs, each neuron processes inputs using a nonlinear activation function enhancing the model's ability to deal with outliers and extreme values. The initial random assignment of weights and biases, followed by iterative adjustments via backpropagation—a method that systematically reduces prediction error—facilitates the learning process. This procedure includes a decay parameter to temper the growth of weights, thus adding a stochastic element to the training process and underscoring the necessity of multiple iterations to achieve consistent and reliable prediction.

For our forecasting exercise, we utilize a simple feed-forward neural network with a single hidden layer, referred to as **Neural Network (NNET)**. This shallow neural network systematically processes information in a sequential manner: starting from the input layer, passing through an intermediary hidden layer, and finally arriving at the output layer. Unlike recurrent neural networks, this architecture ensures a one-way flow of data without any feedback loops, where each layer's output becomes the next layer's input. An example in Figure 2 shows neural net configuration with four inputs and five neurons within the single hidden layer. The hidden layer applies nonlinear activation functions, such as the sigmoid function, to the inputs, enabling the model to learn complex relationships between the predictors and the target variable.

Figure 2: Forward-feed Neural Network with Four Inputs, Five Neurons in One Hidden Layer



Adopting the notation from Hyndman and Athanasopoulos (2018), we sketch the computational process within the neural network. For example, the inputs into hidden neuron *j* are are first linearly combined as shown:

$$z_j = b_j + \sum_{i=1}^{4} w_{i,j} x_i \tag{10}$$

where represents b_j the bias for neuron j, enhancing the flexibility of the model, and $w_{i,j}$ are the weights applied to inputs x_i, directing the strength of connections. This linear combination z_j is then transformed by a nonlinear sigmoid function:

$$s(z) = \frac{1}{1+e^{-z}} \tag{11}$$

facilitating the model's ability to capture complex, nonlinear relationships. The bias and weight parameters b_1, b_2, b_3, b_4, b_5 and $w_{1,1}, \ldots, w_{4,5}$ are optimized from the data, aiming to minimize the network's overall prediction error. o control the magnitude of weights and mitigate the risk of overfitting, a decay parameter - commonly set around 0.1 - is employed, adjusting the weights' update during training. Initially assigned at random, these weights introduce variability to the model's predictions. Consequently, neural networks undergo multiple training iterations with varied initial weights, with the final prediction achieved by averaging these outcomes.

A particular case within feed-forward networks with a single hidden layer is the **Neural Network Autoregressive Model (NNETAR)**, which incorporate lagged values of the target variable as inputs. This method mirrors the linear autoregression approach but introduces the capacity to model complex nonlinear dependencies and seasonal patterns without the constraints for data stationarity. NNETAR models, therefore, represent a fusion of traditional AR behaviors with the capabilities of neural networks' hidden layers to process and forecast based on non-linear relationships. A key aspect of optimizing NNETAR models lies in determining the appropriate number of neurons in the hidden layer and the length of the autoregressive term, which usually is selected based on the optimal number of lags as indicated by the Akaike Information Criterion (AIC) for non-seasonal time series.

Expanding on this concept, we also explore the **Neural Network Autoregressive Model with External Regressors (NNETARX)**, analogous to the ARMAX model. NNETARX models incorporate external or exogenous variables into the forecasting framework, offering an enriched model capable of accommodating additional variables that influence the target variable, similar to the way ARMAX extends the capabilities of simple AR models.

The meanfulness of transitioning to more complex neural networks, such as deep learning models, for certain tasks as inflation forecasting remains a subject of ongoing debate within the field. Empirical evidence from several studies has underscored the advantages of simple neural networks. For instance, simple neural networks have been demonstrated to be the preferred choice in the works of Nakamura (2005) and Araujo and Gaglianone (2022). One possible explanation for the observed superiority of simpler models lies in their ability to effectively capture underlying patterns without the need for extensive data sets typically required by deep learning approaches.

3.8 Factor Model and Model Averaging

Using **Factor Models** to forecast inflation has a long tradition in economic literature, starting from seminal work of Stock and Watson (2002). It proposes a solution to the challenge of forecasting a single series where the predictors outnumber the time series observations. It presents a simplified approach through the use of unobserved latent factors for modeling covariability, employing a two-step forecasting process that involves estimating these factors from the predictors via principal components and then establishing their linear relationship with the forecast variable, keeping consistency even with correlated idiosyncratic errors. More formally, let $x_{t,i}$ for $i = 1, ...N$ be a part of N-dimensional set of potential predictors to forecast y_t. Then the factor representation of data is given by:

$$x_{t,i} = \lambda_i' F_t + e_{i,t} \tag{12}$$

where F_t is a vector of common factors, λ_i is a vector of factor loadings associated with a common factors and $e_{i,t}$ ia an indiosyncratic component of data. As common factors and their loadings are not directly observable, they are estimated using principal component analysis on data, as they are consistent estimators of true latent factors, as shown in Stock and Watson (2002). After extracting common factors, whose number is determined by Bai and Ng (2002), we forecast using estimation on y_h in the direct way:

$$y_t = \beta_F F_t + e_t \tag{13}$$

Bayesian Model Averaging (BMA) has gained recognition as a pivotal approach in linear regression, particularly when confronting scenarios with a large pool of potential regressors against a backdrop of relatively limited observations. At its core, BMA distinguishes itself by incorporating model uncertainty into both predictive and inferential tasks, utilizing posterior probabilities to weigh models proportionately to their evidential support. This technique, based in the principles of Bayes' theorem, advocates for a probabilistic and integrative treatment of the model space M rather than depending on a single model selection process.

The effectiveness of BMA in enhancing predictions of consumer inflation has been substantiated by numerous studies, such as those by Wright (2009), Jacobson and Karlsson (2004), and within a dynamic framework by Koop and Korobilis (2012). In contexts of model uncertainty, BMA transitions from a deterministic to a probabilistic approach, as detailed by Benecká and Komarek (2018), by considering all possible models $M_j, j = 1, ...J$ in the model space M. We need to specify a prior $P(M_j)$, which given data y will lead to posterior probability $P(M_j|y)$. The highest posterior model probability suggests the "best" model given a certain model space and underlying data, while posterior inclusion probability indicates the percentage of times that the variable appears in the models, i.e. variable importance. For technical details, see Wright (2009), among others.

3.9 Hybrid Models

The Random Forest method has advantages in both selecting relevant variables and capturing nonlinear relationships within datasets. To further explore and quantify the relative importance of these two aspects, we employ **Hybrid RF-OLS model**, presented in Medeiros et al. (2021). This model synergizes the variable selection capability of RF with the linear modeling strengths of Ordinary Least Squares (OLS) regression. Specifically, we initiate the process by generating 500 bootstrap samples to grow single trees with a fixed size of *k=25* nodes each. This step is crucial for identifying the variables that RF consistently selects for splitting, thereby highlighting their significance. Subsequently, an OLS regression is applied exclusively on these identified variables to compute forecasts for the dependent variable, with the final forecast derived as the average across all bootstrap samples.

This methodological framework allows for a direct comparison between the predictive performances of the RF and the hybrid RF-OLS models. Should the forecasts produced by both models converge closely, it would suggest that RF's superiority predominantly stems from its adept variable selection. Conversely, a significant divergence in predictive performance would indicate that RF's ability to model non-linear relationships plays a critical role.

In response to the complex task of selecting optimal models for each PPI subsector — which exhibit different pricing behaviors and dependencies — we introduce a suite of hybrid models as a complete innovation. Our motivation stems from the need to accommodate the diverse autoregressive characteristics and sensitivities to input and commodity prices across sectors, without the impracticality of manually fine-tuning models.

Table 1: List of Methods

Type of method	Method	Description
Benchmark	RW	Random Walk
AR-based	AR	Autoregressive Model
AR-based	ARMA	Autoregressive Moving Average
AR-based	ARMAX	ARMA with Exogeneous Variables
Penalized regression	RIDGE	Ridge Regression
Penalized regression	LASSO	Lasso
Penalized regression	ELNET	Elastic Net
Penalized regression	ADALASSO	Adaptive LASSO
Penalized regression	SVR	Support Vector Regression
Tree-based	RF	Random Forest
Tree-based	QRF	Quantile Regression Forest
Tree-based	XGB	Extreme Gradient Boosting
Tree-based	CATBOOST	Gradient Boosting on Decision Trees
Neural nets	NNET	Neural Network
Neural nets	NNETAR	Neural Network Autoregressive Model
Neural nets	NNETARX	NNETAR with Exogeneous Variables
Factor/BMA	FC	Factor Model
Factor/BMA	BMA	Bayesian Model Averaging
Hybrid OLS	RF-OLS	Random Forest with OLS
Hybrid OLS	XGB-OLS	Extreme Gradient Boosting with OLS
Hybrid OLS	CTB-OLS	Gradient Boosting on Decision Trees with OLS
Hybrid ARMAX	RF-ARMAX	Random Forest with ARMAX
Hybrid ARMAX	XGB-ARMAX	Extreme Gradient Boosting with ARMAX
Hybrid NETS	RF-NNETARX	Random Forest with Neural Nets
Hybrid NETS	XGB-NNETARX	Gradient Boosting with Neural Nets

Hence we propose leveraging the variable selection capabilities of tree-based models (Random Forest, XGBOOST, and CatBoost) and integrating these findings with traditional forecasting methods such as Ordinary Least Squares, Autoregressive Moving Average and Neural Network Autoregression, both with external regressors. This method utilizes the unique identification schemes of tree-based models to capture a broad set of potentially influential variables, selecting those that contribute at least 1% to cumulative gain in boosting models or decrease in node impurity in Random Forest, which are our measures of variable importance. The models we propose are as follows: **XGB-OLS, CTB-OLS, RF-ARMAX, XGB-ARMAX, RF-NNETARX, XGB-NNETARX**. Each hybrid model utilizes variable selection insights from tree-based algorithms—namely XGBOOST, CatBoost, and Random Forest —and integrates these with the forecasting precision of Ordinary Least Squares (OLS), Autoregressive Moving Average (ARMA), and Neural Network Autoregressive Models with external regressors (NNETARX).

In developing our hybrid models we leverage the advanced predictive capabilities of ML methods to enhance forecasting accuracy. However, as noted by Berk (2008), caution must be exercised when using ML-derived variable importance measures for variable selection, due to potential issues like overfitting and reduced interpretability. Random Forest approach, for instance, is not inherently designed for variable selection, and its importance measures can be affected by factors such as multicollinearity and the number of predictor levels. To address these concerns within our hybrid approach, we implement cross-validation and limit model complexity to mitigate overfitting risks. Additionally, we interpret the variable importance rankings with care, acknowledging that they reflect predictive strength rather than causal influence.

To summarise, we aim evaluate a comprehensive set of 23 models to determine their forecasting capabilities compared to Random Walk and ARMA. Unlike previous studies, our objective is not to demonstrate supremacy of machine learning methods over Autoregressive Model, but to setup a flexible and practical suite of modelling choices tailored to our task, so the key benchmark is naive. Still, we aim to discuss the forecasting strength of other methods and stay in line with the current literature, so we consider ARMA model as a good start for model performance comparison.

We start with 3 AR-based models and 5 penalized regression methods and similar. Than we investigate forecasting properties of approaches based on trees and neural nets, three of them each. Additionally, we assess the forecasting potential of a factor model and a model utilizing Bayesian model averaging, each offering unique approaches to synthesizing information and accounting for uncertainty in predictions. Finally, we look at 7 hybrid models - a combination of tree-based variable selection with the predictive modeling of autoregressive approaches. An overview of these models, along with the notation adopted for this study, is systematically presented in Table 1.

4. Estimation and Results

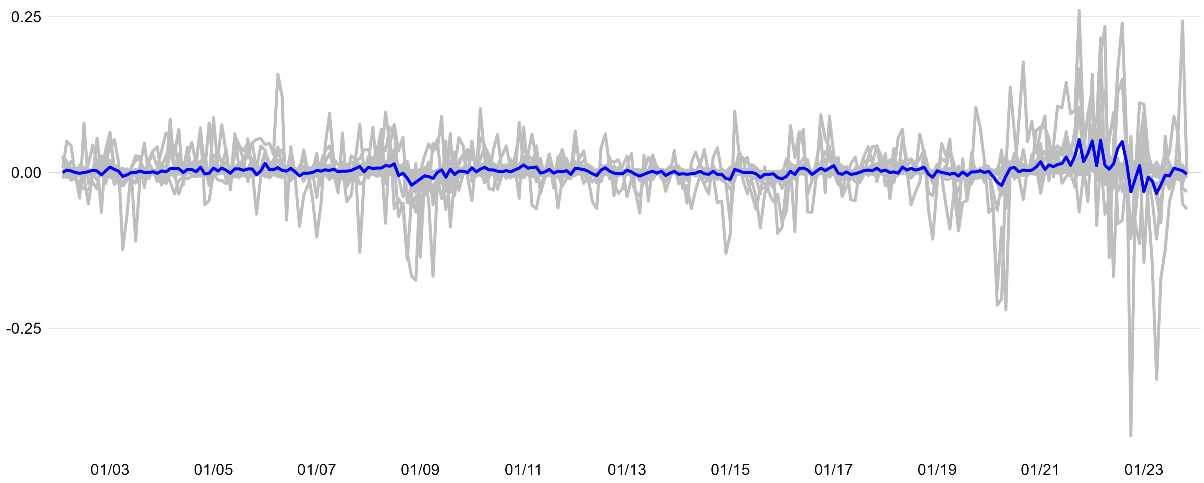
4.1 Data and Estimation Setup

In this paper, we concentrate on producer inflation as measured by changes in the Producer Price Index (PPI). At the disaggregated level, we utilize 29 subgroups according to European Classification of Economic Activities (NACE Rev. 2) at the 2-digit level from the Eurostat database on producer prices, as detailed in Table A.2. Notably, a headline PPI is derived from aggregation using fixed weights from 2015.[5] Eurostat employs a dynamic weighting scheme based on the total turnover value in the respective industry to compile the total PPI, with weights updated every five years. Although changes in the weighting scheme are generally minimal, for simplicity, we have adopted the latest weights available, which cover the period from the year 2000 onward. Unfortunately, data on prices in the C33 group (Repair and installation of machinery and equipment), which accounts for 2% of the aggregate PPI, are only available starting from 2010 and are therefore excluded from this study. The weights for the other subgroups have been adjusted accordingly.

Producer price inflation is expressed as the logarithmic difference in the Producer Price Index. Figure 3 illustrates the evolution of this indicator, both in total (blue) and at the disaggregated level (gray). During most of the period under review, total producer inflation remained within a typical range for month-on-month changes, oscillating between -1% and +1%. However, there were periods of notable fluctuations: a dramatic decline during the global financial crisis and at the onset of the pandemic, and significant increases during energy crises in Europe. At the disaggregated level,

[5] For weights see `https://circabc.europa.eu/ui/group/c4b1ed1e-71ed-4622-b11e-d91429dd96f4/library/d72689ec-103e-41a8-81d1-2e5ea5f171f5?p=1&n=10&sort=modified_DESC`

Figure 3: Euro Area PPI, Log One-month Differences, Total (blue) or 2-digit Subsectors (gray)



Source: Eurostat, author's calculations

we observe that while downswings and upswings can be synchronized, the impact of sector-specific shocks is pronounced. Furthermore, volatility varies significantly across sectors. For instance, prices in sectors like Printing or Pharmaceuticals tend to follow a steady path. Conversely, energy sectors such as the Manufacture of Coke and Refined Petroleum products, and more recently, Electricity and Gas, have experienced considerable volatility due to recent energy crises. These observations align with previous studies that analyze pricing policies based on firm-level data (Gautier, 2008).

As for the drivers of PPI dynamics, we incorporate a set of 24 monthly indicators encompassing commodity prices, costs, production, and sentiment. Alongside price indices for individual commodities such as Brent oil and natural gas, we also include aggregate commodity indices for food and metals. The primary data source for these commodity prices is Bloomberg, with prices converted into euros using the official USD/EUR exchange rate provided by the European Central Bank. Additional data on industrial production and sentiment, categorized under NACE Rev. 2 at the 2-digit level, are sourced from Eurostat. Sentiment indicators, namely the Economic Sentiment Indicators (ESI) from the European Commission, are presented as both monthly and quarterly balances of survey responses. Monthly balances include responses regarding orders, production, selling price expectations, and stocks from manufacturing firms, while quarterly balances focus on factors limiting production such as labor, demand, and material and equipment availability[6]. We also incorporate time series on factors limiting production, which effectively capture the recent intensification of global value chains and the corresponding rise in input costs. For similar reasons, we include the cost of shipping goods worldwide, as indicated by the Baltic Exchange Dry Index from Bloomberg, and the Global Supply Chain Pressure Index (GSCPI) from the New York Fed.[7]

[6] Quarterly data are converted to monthly indicators using methods of temporal disaggregation, specifically the Denton-Cholette method. Refer to the R package *tempdisagg* for practical implementation details and references. As sentiment indicators are only available for manufacturing groups (C group), we utilize total indices for energy sectors (B, E, and D groups) to reflect the broader economic environment.
[7] `https://www.newyorkfed.org/research/policy/gscpi#/interactive`

Table A.1 in Appendix A.1 provides a comprehensive list of variables categorized by their source. Data on producer prices are seasonally adjusted using the X13 method, and some series from Eurostat are obtained as directly seasonally adjusted data. To ensure stationarity, individual time series transformations are applied according to standard practices in the literature. Specifically, PPI, industrial production, and commodity/transportation prices are transformed using log differences, while sentiment indicators such as the Economic Sentiment Indicator or the Global Supply Chain Pressure Index are presented in simple differences. A maximum of two lags are considered for each series.

Stationarity of the transformed data series is verified using the Phillips–Perron test at a 5% significance level. This choice of test is motivated by the presence of substantial structural changes, particularly in the energy markets over the sample period.[8] Notably, energy prices have become more market-driven. There has been a significant shift towards renewable energy sources, and long-term contracts for non-carbon neutral energy sources have become less attractive due to new environmental policies, such as emission allowances and carbon taxes in EU countries. These market dynamics have inevitably influenced pricing policies in production sectors. Particularly, natural gas has been considered an optimal transitional energy source. However, its security was compromised following the Russian invasion of Ukraine and the subsequent cessation of Russian gas deliveries to Europe. This disruption led to periods of increased volatility in commodity prices, which, in turn, spilled over to producer prices. The post-pandemic reopening of economies exerted additional pressure on supply chains, leading to rising prices across various inputs, including transportation. Furthermore, input price shocks were often accompanied by profit-price spiral tendencies, as documented in several analytical studies.[9]

The dataset spans from February 2002 to November 2023, providing a total of 251 observations for each variable. The model estimation sample extends from February 2002 through June 2019, comprising 209 observations. We consider forecast horizons (h) of 1, 3, and 6 months. For $h=1$, the evaluation period begins in July 2019, yielding 53 forecasts, while for $h=6$, it starts from December 2019, resulting in 48 forecasts. Time series cross-validation is conducted on data prior to the onset of the pandemic (i.e., until February 2020) and on the complete dataset to assess how preferred parameters have adapted in response to the increased volatility post-pandemic.

The models are re-estimated monthly in a recursive manner, expanding by one observation each time. We adopted this recursive approach due to its efficiency in our context. Alternatively, Medeiros et al. (2021) employs a rolling-window scheme to accommodate potential structural breaks in their models aimed at forecasting U.S. inflation beginning in the 1960s. Given that our sample period is considerably shorter, employing a rolling-window method in our study would pose challenges due to the limited sample size.

For parameter selection, we adhere to best practices as documented in the relevant literature, see for example Medeiros et al. (2021) or Schnorrenberger and Moura (2024). In our analysis, the penalty parameter λ for shrinkage methods is determined using the Bayesian Information Criterion (BIC), following the recommendations by Araujo and Gaglianone (2022). For the Elastic Net penalty, the α parameter is set at 0.5. The results obtained using BIC as the selection criterion show consistent outcomes. For the Adaptive LASSO, the weights are derived from the β coefficients of the non-adaptive version of the model. In terms of Random Forest, each tree is allowed to grow until each leaf (terminal node) contains only five observations. The number of variables considered for

[8] For an overview of energy market legislation, see `https://www.europarl.europa.eu/factsheets/en/sheet/45/internal-energy-market`

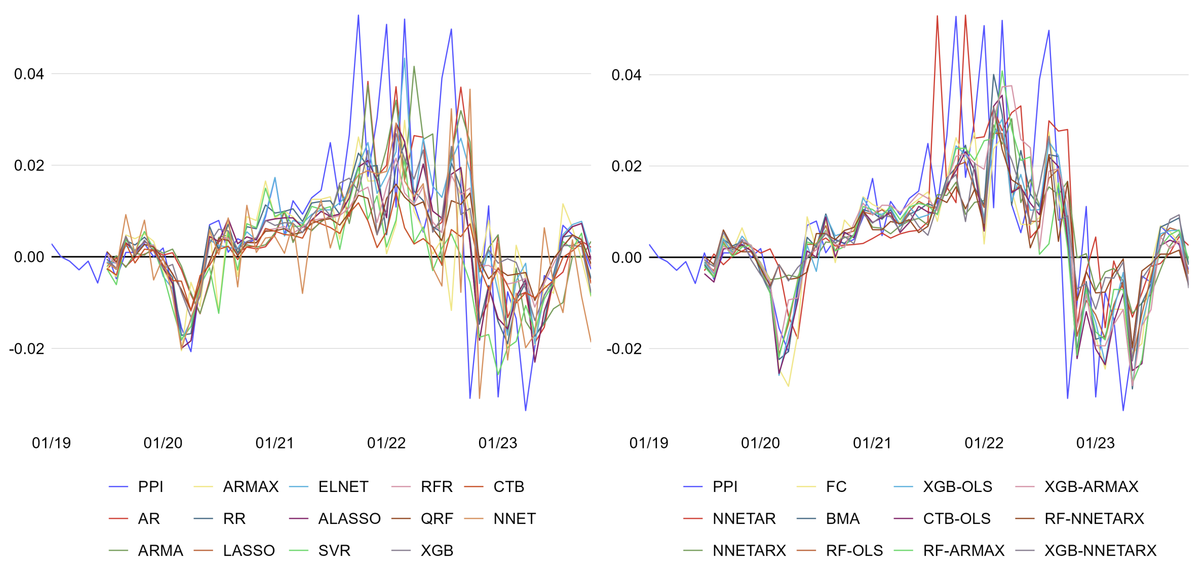
[9] For analysis on euro area producer price data see GEV-2022 or ECB-2021

splitting at each node is fixed at 39. The neural network is configured simply, with a linear output activation function suitable for regression tasks and 10 units (neurons) in the hidden layer. The Support Vector Regression employs a linear kernel with its cost parameter set to 0.25. For Bayesian Model Averaging, unit information priors are used for parameters, and a uniform prior is applied to the model space, indicating no initial preference among models. The top 5 models are selected based on the Occam's window criterion. Detailed discussions on the impact of these parameter settings on our forecasting exercise are provided in Appendix B.

4.2 Results

To illustrate model performance, Figure 4 displays the observed logarithmic differences in the total PPI against its 1-month ahead out-of-sample forecasts generated by all models discussed in this paper.[10] An initial review indicates that most methods performed effectively during the pre-pandemic period (i.e., until early 2020). This observation will be considered in further analyses. Model performance begins to deteriorate from late 2021 onwards, coinciding with post-pandemic economic reopening, which intensified pipeline pressures and input prices. Models also struggle to capture the abrupt price fluctuations starting in early 2022, triggered by the Russian invasion of Ukraine, which precipitated an unprecedented energy crisis in Europe, significantly impacting energy prices. However, the overall performance appears to improve in the second half of 2023 as the effects of earlier shocks diminishes.

Figure 4: Actual vs. Estimated Log One-month Differences in Total PPI



For each method and forecast horizon, we compute the out-of-sample root mean squared errors. Figure 5 illustrates the RMSE levels across models and sectors for a 1-month ahead horizon for illustration. Given the varied results across sectors, the sample is divided into three subsamples with similar outcomes. The highest RMSE values are typically found in sectors closely linked to oil prices, such as B06 - Extraction of crude petroleum and natural gas, and C19 - Manufacture of coke and refined petroleum products. These are followed by sector D35 - Electricity, gas, steam, and

[10] Results for individual subsectors as well as aggregate totals across all forecast horizons are available from the author upon request.

air conditioning supply, which reflects the price fluctuations of gas and electricity in the production sector. The disparity between the performances of the models is notably wide in these sectors. Only two additional sectors, B07 - Mining of metal ores and B05 - Mining of coal and lignite, exhibit higher average RMSEs than the total PPI. Sectors whose pricing policies are closely tied to volatile energy commodity prices generally exhibit poorer performance. Conversely, the lowest RMSEs are found in sectors with higher value-added, where pricing policy does not adhere to a straightforward rule. This includes machinery-related groups such as C28 - Manufacture of machinery and equipment and C30 - Manufacture of other transport equipment, as well as C21 - Manufacture of basic pharmaceutical products. The differences in pricing policies, as documented by Gautier (2008), are also observable in our analysis, reflecting the unique dynamics within these sectors.

Figure 5: Root Mean Squared (RMSE) Errors Across Sectors and Methods for 1 Month Ahead Forecasts

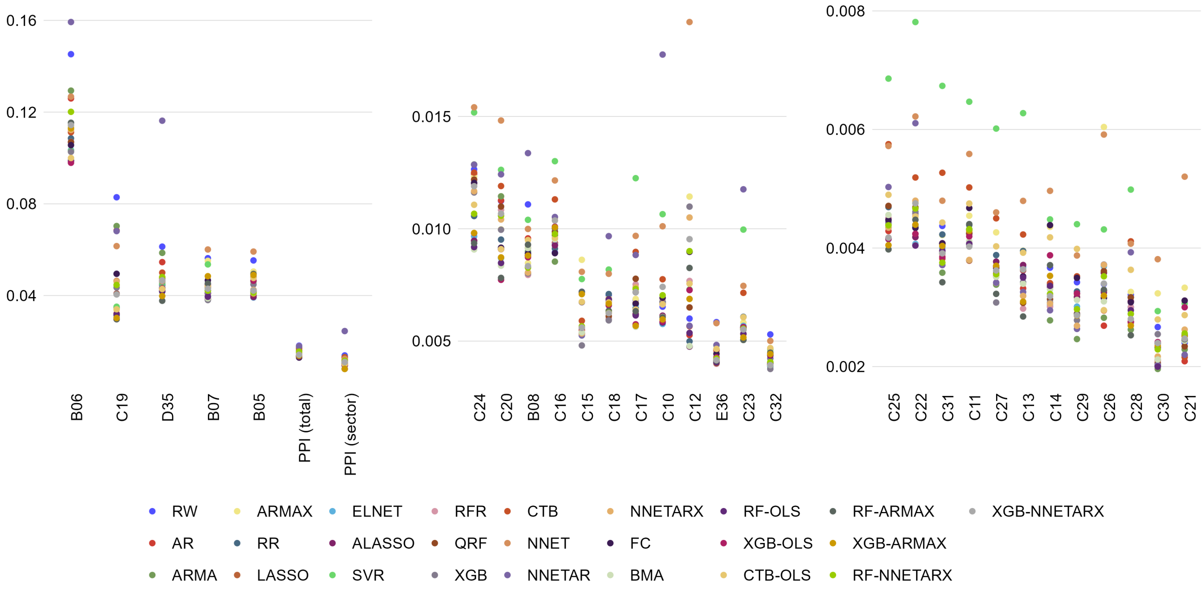


Table 2 presents the relative RMSE for 1-month ahead forecasts across all sectors and methods, with the ARMA model serving as the benchmark. Here, values less than one indicate superior performance relative to the benchmark. The results for the 3-month and 6-month horizons are displayed in Tables C.1 and C.2 in Appendix C, respectively.[11] In these tables, the three lowest RMSE values for each sector are highlighted in red. An asterisk indicates the p-value from the Diebold and Mariano (2002) test, which tests the null hypothesis of no difference in the accuracy of two competing forecasts. Standard notation is used to denote levels of statistical significance: * for the 10% level, ** for the 5% level, and *** for the 1% level, for any table forward.

A preliminary review of our results underscores the success of our strategy to utilize a diverse array of models. It is evident that no single method consistently outperforms others across all sectors simultaneously. As previously emphasized, production sectors in the euro area exhibit distinctly different pricing policies, necessitating a model-intensive approach. Our analytical framework is

[11] Our primary focus is on forecasting results at the shortest horizon, given the limited availability of forecasts for explanatory variables over longer periods. The advantage of leveraging a data-rich environment is predominantly evident in nowcasting. Nonetheless, to maintain alignment with existing literature, we will also report results for longer horizons.

structured such that for each subsector, even at the shortest forecast horizon, we identified a model that achieved a lower RMSE than a Random Walk.

At the shortest forecast horizon, it is often challenging to outperform our benchmark, the ARMA model, which itself sometimes fails to surpass a simple Random Walk. Among the penalized regression methods, the Elastic Net consistently shows the best performance. However, Support Vector Regression generally does not outperform the ARMA benchmark in most cases. Tree-based methods, particularly Random Forest Regression and Quantile Random Forest, perform comparably well, with XGBoost excelling in a number of sectors, whereas CatBoost is less successful. Among neural network approaches, only NNETAR shows some success in predicting sectoral PPI, yet it still falls short of surpassing the ARMA benchmark. Using factor models appears reasonable for forecasting total PPI, but they have limited effectiveness for individual sectors. Bayesian Model Averaging performs robustly, as its RMSE values are lower than those of ARMA in numerous sectors, often at a statistically significant level. Hybrid methods, specifically RF-OLS and XGB-OLS, outperform our benchmark in several cases, while CTB-OLS generally performs poorly. Hybrid ARMAX models, particularly RF-ARMAX, are typically the most successful, consistently ranking among the top three models for a given sector. XGB-ARMAX is slightly less effective, whereas Hybrid NETS lag behind their peers in most instances.

To summarize, on a one-month horizon, penalized regression methods (excluding SVR) and tree-based methods demonstrate some forecasting efficacy, albeit not significantly stronger than that of the ARMA model. BMA and RF-OLS appear to be viable choices, while factor model is predominantly useful for forecasting aggregates. The hybrid RF-ARMAX model stands out as a robust option for forecasting short-term PPI inflation across sectors, with XGB-ARMAX also presenting a solid alternative. In contrast, methods based on neural networks or CTB substantially lag behind.

On longer forecast horizons, the efficacy of our alternative approaches increases, and their predictions achieve greater statistical significance compared to the ARMA model. At the three-month horizon, the Hybrid RF/XGB-OLS and Bayesian Model Averaging excel, with Hybrid RF/XGB-ARMAX and penalized regression methods also performing well in some cases. Even more substantial gains in forecasting accuracy are observed at the six-month horizon, where RF/XGB-OLS and BMA emerge as leaders. Additionally, tree-based methods, Hybrid RF/XGB-ARMAX, and penalized regression methods now show stronger results. Hybrid NETS and NNETARX can now surpass our ARMA benchmark, although NNET still underperforms. These results are consistent with the literature, which indicates that the advantages of non-AR-based methods are typically more pronounced at longer horizons. Furthermore, a significant portion of our forecasting exercise's sample was characterized by unexpected increases in commodity prices, transport costs, and margins. Incorporating this information through alternative methods has significantly enhanced forecast accuracy.

Upon analyzing sector-specific performance, it is evident that even at the shortest forecast horizon, our alternative methods successfully predict trends in energy-related sectors. These include C19 - Manufacture of coke and refined petroleum products, C20 - Manufacture of chemicals and chemical products, and to a lesser extent, D35 - Electricity, gas, steam, and air conditioning supply. Together, these sectors constitute nearly one-third of the total Producer Price Index. Most models under consideration demonstrate robust forecasting capabilities for these sectors, accurately identifying the primary driving forces behind their price movements. The RF-ARMAX method stands out particularly in forecasting these groups, leveraging the strength of Random Forest to pinpoint key drivers and integrate these dynamics with simple autoregressive price behavior.

Table 2: Relative RMSEs for One-month-ahead Forecasts—Baseline

		AR-based			Penalized regression						Tree-based			Neural nets			Factor/BMA			Hybrid OLS		Hybrid ARMAX		Hybrid NETS	
Sector	RW	AR	ARMA	ARMAX	RR	LASSO	ELNET	ALASSO	SVR	RFR	QRF	XGB	CTB	NNET	NNETAR	NNETARX	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX	RF-NNETARX	XGB-NNETARX
PPI (total)	1.05	0.97	1	0.94	0.77**	0.79**	0.78***	0.79**	0.89	0.88**	0.92	0.85**	0.99	1.03	1.02	0.87**	0.75**	0.79**	0.78**	0.78**	0.8**	0.79**	0.8***	0.9*	0.82**
PPI (sector)	1.06	0.99	1	0.77**	0.78**	0.78**	0.75***	0.78**	0.92	0.85**	0.89	0.82**	0.99	0.94	1.86	0.84**	0.73**	0.75**	0.73**	0.74**	0.77**	0.61***	0.61***	0.88	0.84*
B05	1.33	0.96	1	1.21	0.96	0.94*	0.94*	0.94*	1.11	1	0.99	1.08	0.98	1.42	1.02	1.02	1.15	0.99	1	1.12	1.14	1.19	1.17	0.99	1.01
B06	1.12	0.97	1	0.88	0.84**	0.76**	0.8***	0.76**	0.8**	0.83**	0.83**	0.79***	0.86**	0.98	1.23	0.87**	0.82**	0.76**	0.76**	0.76**	0.77**	0.89	0.87	0.93	0.88*
B07	1.3	1.01	1	1.27	0.97	0.89*	0.92	0.89*	1.24	0.94	0.95	0.88*	0.94	1.39	0.95	1	1.07	0.89	0.91	1.03	1.02	1.04	1.12	0.97	0.99
B08	1.23	1.06	1	0.94	0.99	0.98	0.92	0.98	1.16	0.88	0.89	0.89	1.02	1.11	1.49	0.9*	0.99	1.02	0.98	0.97	1.06	1.03	0.98	0.92	0.93
C10	0.98	1	1	1	1.01	0.92	0.86	0.92	1.6	0.99	1.01	1.01	1.16	1.52	2.66	1.04	1.04	0.91	0.87*	0.88	0.99	0.91	0.89	1.06	1.11
C11	1.14	1	1	1.2	1.16	1.12	1.12	1.12	1.7	1.08	1.12	1.07	1.32	1.47	1.14	1	1.23	1.08	1.07	1.11	1.25	1.16	1.13	1.14	1.06
C12	1.26	1.11	1	2.4	1.05	1	1	1	1.53	1.61	1.37	2.31	1.19	4.04	1.19	2.21	1.89	1.01	1.13	1.53	1.59	1.74	1.45	1.89	2
C13	1.15	1.04	1	1.24	1.24	1.16	1.15	1.16	1.96	0.93*	0.96	1.07	1.32	1.5	1.02	1	1.1	1.06	1.1	1.13	1.23	0.89**	0.97	1.14	1.14
C14	1.32	1.08	1	1.57	1.22	1.2	1.09	1.2	1.61	1.09	1.13	1.11	1.23	1.79	1.06	1.1	1.58	1.19	1.21	1.39	1.5	1.34	1.27	1.16	1.15
C15	1.34	1.02	1	1.61	1	1	0.98	1	1.45	1	0.99	0.9**	1.1	1.51	0.98	1	1.26	1.01	1.04	1.34	1.26	1.34	1.32	1.06	1.04
C16	1.05	1.1	1	1.23	1.07	1.08	1.09	1.08	1.52	1.15	1.16	1.17	1.32	1.42	1.23	1.12	1.04	1.12	1.14	1.18	1.12	1.17	1.18	1.14	1.21
C17	1.02	1	1	1.03	0.97	0.93	0.99	0.93	1.84	1.13	1.16	1.07	1.34	1.45	1.32	1.11	1	0.84**	0.92	0.86**	0.96	0.95	0.85**	1.1	1.08
C18	1.18	1.05	1	1.16	1.19	1.15	1.1	1.15	1.37	1.01	1.02	0.99	1.14	1.34	1.62	1.06	1.13	1.14	1.15	1.11	1.13	1.06	1.12	1.05	1.05
C19	1.18	0.97*	1	0.48***	0.49***	0.46***	0.43***	0.46***	0.5***	0.58***	0.58***	0.62***	0.66***	0.88	0.97	0.66***	0.7***	0.44***	0.45***	0.42***	0.48***	0.42***	0.43***	0.63***	0.58***
C20	0.98	0.98*	1	0.87	0.83**	0.79***	0.76**	0.79***	1.1	0.95	0.96	0.87*	1.04	1.29	1.09	0.91	0.8**	0.73***	0.74***	0.68***	0.79***	0.68***	0.76***	0.92	0.93
C21	1.05	0.91***	1	1.45	1.06	1.06	1.06	1.06	1.34	1.1	1.02	0.96	0.94*	2.27	0.96*	1.15	1.36	1.06	1.11	1.31	1.25	1.08	1.1	1.12	1.08
C22	1.01	0.97**	1	1.1	1.05	0.96	0.93	0.96	1.79	1.05	1.08	1.06	1.19	1.43	1.4	1.1	1	0.93*	0.93*	0.97	1.04	1.03	1.01	1.07	1.09
C23	1.1	1.05	1	1.15	1.2	1.11	1.12	1.11	1.97	1.09	1.09	1.09	1.41	1.48	2.33	1.11	1.11	1.09	1.05	1.11	1.2	1	1.02	1.1	1.09
C24	0.98	0.97	1	0.96	0.82*	0.74**	0.75**	0.74**	1.18	0.93	0.95	0.9	0.97	1.2	1	0.91	0.94	0.71***	0.71**	0.76**	0.86	0.73**	0.76**	0.83*	0.92
C25	1	0.97	1	1	1.07	1.02	0.99	1.02	1.56	1.02	1.07	1.01	1.31	1.3	1.14	0.99	1.02	1.04	0.94	0.95	1.11	0.9	0.92	1	0.95
C26	1.2	0.95	1	2.14	1.32	1.27	1.31	1.27	1.53	1.32	1.28	1.16	1.04	2.09	1.12	1.32	1.12	1.1	1.15	1.16	1.05	1.17	1.13	1.25	1.2
C27	1.09	1.04	1	1.26	1.15	1.11	1.07	1.11	1.78	1.01	1.06	0.91	1.33	1.36	1.01	1.04	1.08	1.04	1.09	1.09	1.19	0.95	1.09	1.05	1.07
C28	1.22	1.05	1	1.24	1.21	1.18	1.14	1.18	1.9	1.15	1.21	1.05	1.57	1.55	1.5	1.04	1.18	1.1	1.12	1.05	1.38	0.96	1.03	1.1	1.07
C29	1.39	1.17	1	1.3	1.33	1.29	1.22	1.29	1.79	1.09	1.17	1.13	1.43	1.57	1.07	1.09	1.42	1.27	1.31	1.31	1.62	1.17	1.16	1.2	1.16
C30	1.36	1.08	1	1.65	1.02	1.03	1.03	1.03	1.5	1.08	1.05	1.3	1.18	1.94	1.03	1.11	1.22	1.08	1.02	1.23	1.43	1.19	1.2	1.17	1.22
C31	1.22	1.07	1	1.18	1.18	1.14	1.12	1.14	1.88	1.1	1.14	1.09	1.47	1.34	1.04	1.09	1.14	1.09	1.1	1.07	1.24	0.95	1.13	1.05	1.09
C32	1.37	1.02	1	1.18	1.02	1.03	1	1.03	1.2	1.05	1.04	0.97	1.15	1.3	1.09	1.03	1.13	1.02	1.04	1.13	1.21	1.16	1.14	1.05	1.02
D35	1.05	0.93	1	0.75*	0.76**	0.74**	0.74**	0.74**	0.77*	0.79**	0.81*	0.79**	0.85	0.82	1.98	0.79**	0.71**	0.73**	0.72**	0.74**	0.73**	0.64***	0.68**	0.82*	0.8**
E36	1.41	1.11	1	1.07	0.98	0.98	0.98	0.98	1.16	0.96	0.97	0.99	1.12	1.4	1.17	0.99	1.07	0.98	0.98	1.03	1.12	1	1.01	1.02	1

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: RW - Random Walk, AR - Autoregressive Model, ARMA - Autoregressive Moving Average, ARMAX - ARMA with Exogeneous Variables, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, ALASSO - Adaptive LASSO, SVR - Support Vector Regression, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, CATBOOST - Gradient Boosting on Decision Trees, NNET - Neural Network, NNETAR - Neural Network Autoregressive Model, NNETARX - NNETAR with Exogeneous Variables, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX, RF-NNETARX - Random Forest with Neural Nets, XGB-NNETARX - Gradient Boosting with Neural Nets

Similarly, RF-ARMAX achieves a lower RMSE compared to ARMA across a set of C sectors (C24 - Manufacture of basic metals, C25 - Manufacture of fabricated metal products, C27 - Manufacture of electrical equipment, C28 - Manufacture of machinery and equipment), although these differences are not statistically significant. This outcome suggests the presence of similar, albeit weaker, pricing behaviors in these sectors. Notably, this pattern persists even on longer horizons, when additional models begin to exhibit improved performance. Section 4.6 offers an explanation of the underlying factors that contribute to the weak performance of our models in certain sectors in the baseline setup.

Penalized regression methods have shown particular success in forecasting most of the B sectors. Conversely, ARMA continues to be among the top three models for numerous high-value-added C manufacturing groups, ranging from C26 - Manufacture of computer, electronic, and optical products to C32 - Other manufacturing. This advantage only marginally improves over longer horizons, suggesting the possibility of omitted variable bias, i.e., that our set of explanatory variables may lack crucial information needed to effectively forecast these sectors.

In Appendix B, we discuss parameter tuning related to our baseline setup. The analysis reveals that while fine-tuning models across various sectors yields some improvement in forecasting accuracy, the overall impact remains modest. The notable exception is Support Vector Regression, which shows significant enhancement from parameter adjustments.

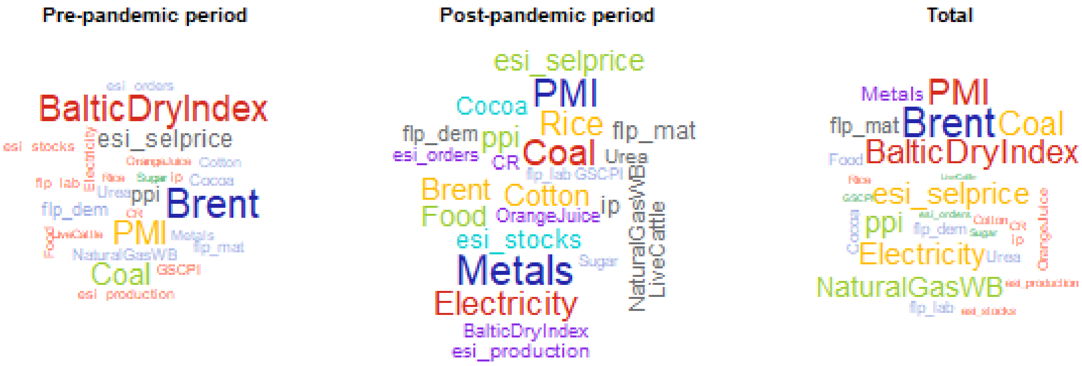
4.3 Variable Selection

Several methods can also provide valuable insights into the driving forces behind price developments. We focus our discussion on the variable importance results from LASSO, Random Forest, XGBoost, and Bayesian Model Averaging, as these methods have shown promising performance in forecasting. We begin with a simple overview of how the importance of different explanatory variables has evolved over time, distinguishing effects between the pre-pandemic period (from the start of our sample until February 2020) and the post-pandemic period (from February 2020 to November 2023). The latter period was characterized by a significant increase in the importance of input variables such as commodity prices and transportation costs, which strongly influenced selling price expectations. While machine learning methods like Random Forest or XGBoost offer valuable insights into the driving forces behind price developments, we interpret their variable importance rankings with caution. Specifically, we acknowledge that these rankings reflect predictive strength rather than causal influence, and thus, we comment on their utility for enhancing forecasting accuracy.

Figure 6 illustrates the results for word clouds, comparing the pre-pandemic and post-pandemic periods to the entire period under consideration. These word clouds are based on the 10 most significant variables identified for each subsector by a selected sample of methods. They effectively display the relative importance of variables across all subsectors, lags, and methods, providing a visual representation of the changing dynamics.

The analysis reveals intuitive patterns. In the pre-pandemic period, the primary input price driver was Brent oil, followed by transportation costs as indicated by the Baltic Dry Index, and coal prices. Economic activity, as measured by the Purchasing Managers' Index (PMI), also had a substantial impact, whereas lags of the PPI and selling price expectations played a relatively minor role. The post-pandemic period has been characterized by a broad-based rise in commodity prices and an energy crisis affecting the prices of electricity and gas. Initially, we might expect the Global Supply Chain Pressure Index to dominate as a measure of supply chain difficulties; however, the data do not

Figure 6: Word Clouds for the Most Important Variables Across Subsectors and Methods



Note: For variable definition, please see Table A.1 in Appendix. The size and colour reflects relative importance of the variable, for details see text.

support this assumption. Instead, aspects of the issue are captured by other measures, such as stock levels (esi_stocks - firms' assessment of finished goods) and transportation costs. Furthermore, the GSCPI may not be the optimal measure for supply chain pressures in the euro area as it reflects global developments in transportation costs and delivery times. A more region-specific measure might be more appropriate, although such metrics are not readily available. For instance, De Santis (2024) utilizes delivery times as a component of PMI to assess supply chain pressure, but these subindices are not available to the Czech National Bank.

Interestingly, several less significant variables, such as Cotton and Rice, unexpectedly increased in importance in post-pandemic period. Two potential explanations emerge for this observation: either the inflation wave was propelled by a broad-based rise in commodity prices affecting all sectors, or the models struggled to identify the best predictors for substantial post-pandemic price increases that were driven more by profit motives than by commodities. Given the substantial body of literature discussed in the first chapter, we are more inclined towards the latter explanation.

Now lets dive deeper into importance across sectors, still keeping a split in pre-pandemic and post-pandemic period. Figure 7 shows the best three variables across sector in pre-pandemic period, with order given by colour - first (blue), second (red) and third (yellow). The chart for post-pandemic period are in Appendix D, see Chart C.1. These visual representations offer intriguing insights, providing substantial evidence about the behavior of individual sectors and the effectiveness of the methods employed. This dual-period analysis highlights the dynamic nature of economic factors influencing sector-specific pricing mechanisms.

Firstly, there is a consensus among models regarding the key drivers of producer inflation in certain sectors, particularly those linked to energy supply or that are energy/commodity intensive. For example, in the Manufacturing of Petroleum Products (C19) and Chemicals (C20) during the pre-pandemic period, the Brent oil price is identified as a critical input and energy component for production. Similarly, sectors like Electricity, Gas, etc. (D35), and Extraction of Crude Petroleum and Natural Gas (B06) heavily depend on the prices of gas, electricity, or Brent oil. For metal-related groups such as Manufacture of Basic Metals (C24), Manufacture of Electrical Equipment (C27), and Mining of Metal Ores (B07), the predominant driver is the price of metals, indicated here by the metal price commodity index.

Second, a number of sectors exhibit a strong reliance on autoregressive behavior for effective inflation forecasting. This is particularly evident in sectors such as C13 (Manufacture of Textiles), C17 (Manufacture of Paper and Paper Products), C31 (Manufacture of Furniture), and C26 (Manufacture of Computer, Electronic, and Optical Products), where past price data significantly influence future trends. A noteworthy exception is found in C10 (Manufacture of Food Products), where lagged values are crucial, yet the food commodity index plays only a marginal role. In this case, one of the most effective models is RF-OLS, which adeptly handles scenarios where lagged behaviors need to be complemented by commodity price inputs. Conversely, C29 (Manufacture of Motor Vehicles, Trailers, and Semi-Trailers) diverges from this autoregressive pattern. Instead, this sector is predominantly influenced by the prices of electricity and transportation costs, which are also indicative of broader supply chain dynamics.

Finally, variability in model selection is prevalent in certain subsectors, such as C11 (Manufacture of Beverages), C14 (Manufacture of Wearing Apparel), C21 (Manufacture of Basic Pharmaceutical Products), C30 (Manufacture of Other Transport Equipment), and E36 (Water Collection, Treatment, and Supply). This variation often arises when models struggle to pinpoint the appropriate explanatory variables, suggesting that multiple factors may be influencing these sectors. For these subsectors, the absence of explicit price indices that are directly relevant poses a significant challenge. Additionally, their pricing policies may be influenced by the broader inflationary environment rather than by their own historical pricing trends. This is particularly evident in sectors like E36, where there could be strong non-linear price patterns due to administered pricing practices, such as annual repricing events typically occurring in January.

As we transition into the post-pandemic period, the clarity of model consensus diminishes. There is considerably less agreement among models concerning both key and supplementary drivers of inflation. Notably, models sometimes select seemingly unrelated variables, such as Cocoa or Cotton in manufacturing, as they attempt to account for unexpectedly high price fluctuations. This observation aligns with our earlier findings from the word cloud analysis, which highlighted an unexpected rise in the importance of variables that are not directly related to the operational dynamics and pricing policies of the sectors.

The results vary also across different methods. Both Random Forest and Boosting Trees utilize tree-based structures, yet they differ significantly in their approaches to measuring variable importance due to their distinct constructions. In the case of Random Forest, variable importance is assessed using Mean Decrease in Accuracy (MDA). After the forest is trained, the values of each feature are permuted one at a time across the out-of-bag samples, and the decrease in model accuracy is recorded. The average decrease in accuracy across all trees when a feature is shuffled indicates the MDA. Thus, a feature that leads to a significant drop in accuracy upon shuffling is considered important. Conversely, XGBoost assesses variable importance through the average gain of splits that utilize the feature. This method aggregates the contribution of the feature across all trees within the model by summing the gains associated with that feature each time it is used in a split. A higher gain indicates that the feature significantly improves the model's performance when used in a split. Therefore, XGBoost has the potential to highlight unexpected features, such as unusual commodity prices (e.g., Cocoa, Orange Juice), particularly in the post-pandemic period. This frameworks might be more difficult to interpret but provides an alternative view that might spurr forecaster's interest to understand driving forces especially in challeging periods.

In the case of Bayesian Model Averaging, we determine variable importance through the posterior inclusion probability across the model space. Unlike methods that depend on a single model's evaluation, BMA mitigates model uncertainty by averaging over a multitude of models, each weighted by their probability of being the most accurate given the data. This method effectively identifies which variables consistently appear in models that are deemed highly probable, thereby providing a probabilistic measure of feature importance. Hence, BMA tends to emphasize variables that are traditionally considered significant, filtering out less relevant data and providing more stable and interpretable results. This approach, while sometimes resulting in a higher error rate during challenging periods, offers a robust framework by integrating model uncertainty into its analysis.

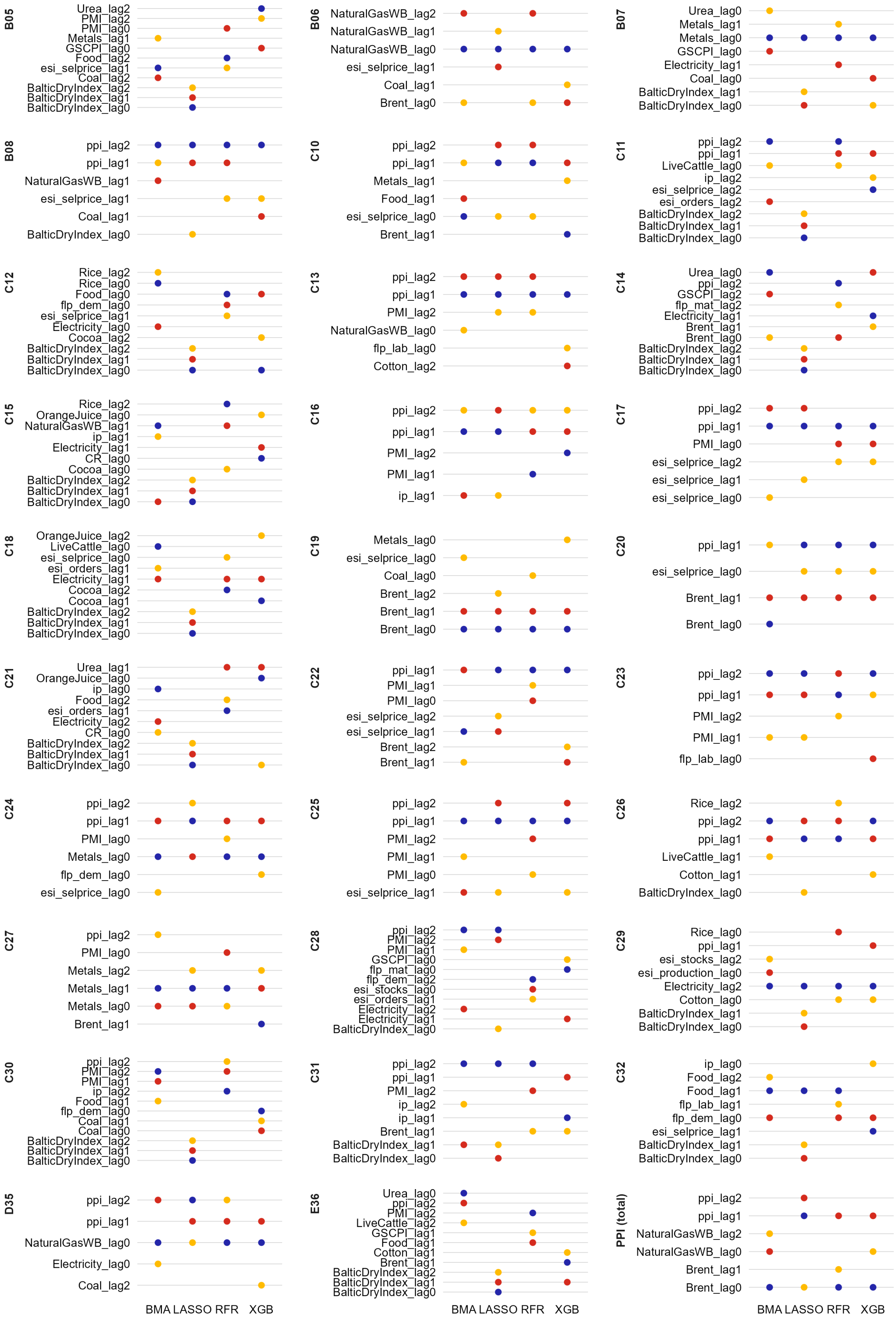
To conclude with LASSO, this method performs variable selection by applying an L1 penalty, which effectively shrinks some of the regression coefficients to zero. The variables that retain non-zero coefficients after the LASSO fitting process are considered significant contributors to the model's predictions. Consequently, LASSO tends to choose only one variable from a group of highly correlated variables, a characteristic driven by the nature of the L1 penalty, that might be difficult to interpret. This is evident in Figure 7, where LASSO consistently selects the Baltic Dry Index as the primary variable for specific sectors (B05, C12, C11, C15, C18, C21, C30, E36), particularly in periods prior to the pandemic. The influence of this feature diminishes during the post-pandemic inflation surge, suggesting that LASSO's strengths are most apparent in more dynamic economic conditions.

4.4 Forecast Combination and Model Selection

As indicated by our baseline results, no single model universally fits forecasting all sectors of the disaggregated producer prices. Thus, it is essential to explore whether any model selection and combination methods can enhance our overall forecasting accuracy. o streamline the model space and focus on the most effective methods, we excluded 11 models from further consideration. This includes the Random Walk model and ADALASSO, whose performance closely mirrors that of standard LASSO estimation. Additionally, we conducted a quantile analysis of root mean squared errors to identify underperforming models. In this analysis, models were ranked based on their RMSE values, and those appearing in the lower quantiles—indicating consistent underperformance relative to others—were considered for exclusion. To ensure we retained a diverse yet efficient model set, we applied a 40% exclusion threshold, balancing model variety with practical forecasting performance across different horizons and sectors. Consequently, the nine least effective models were removed: ARMAX, SVR, CTB, NNET, NNETAR, NNETARX, CTB-OLS, RF-NNETARX, and XGB-NNETARX. These models also failed to rank among the top three based on RMSE compared to ARMA at the 1-month horizon, as highlighted in red in Table 2.

For forecast combination and model selection purposes, we now focus on the remaining 15 models (AR, ARMA, RR, LASSO, ELNET, RFR, QRF, XGB, FC, BMA, RF-OLS, XGB-OLS, CTB-OLS, RF-ARMAX, XGB-ARMAX). These models represent a diverse set that has shown relative strength in our baseline analysis and offers a comprehensive approach to modeling the varied dynamics of producer price inflation.

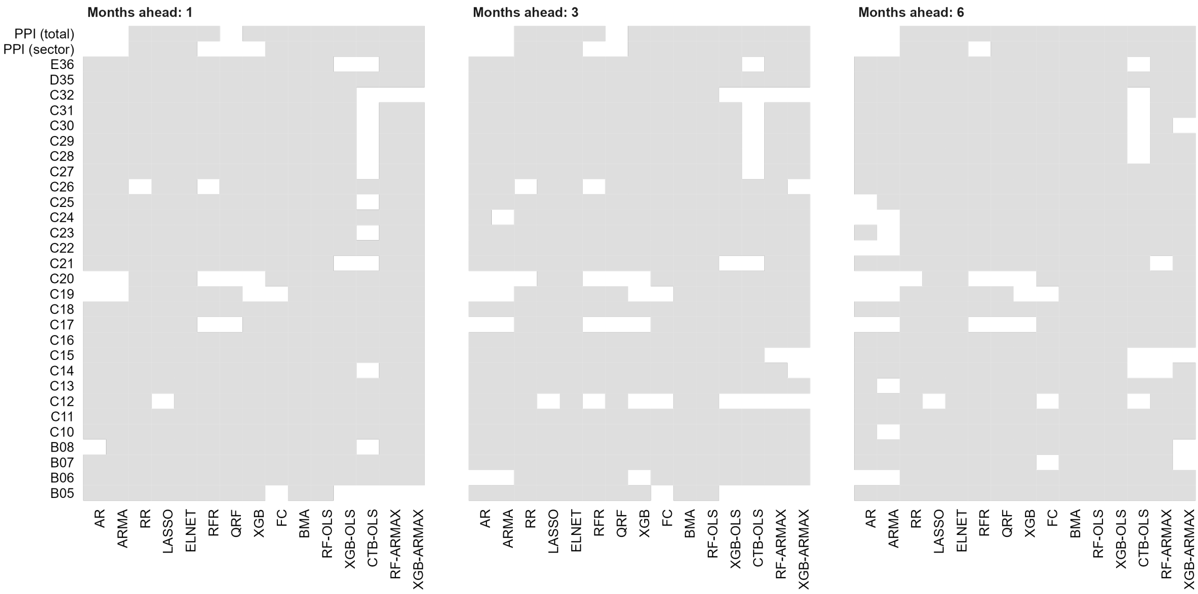
Figure 7: Ranking of the Three Best Variables (Pre-pandemic Period)



Note: Sector code is on left, estimates for pre-pandemic period, order given by colour - first (blue), second (red) and third (yellow).

The first model selection set is based on **Model Confidence Set (MCS models)** approach, introduced in Hansen et al. (2011). The MCS procedure employs a sequence of statistical tests designed to construct a group of "superior" models, where the null hypothesis of Equal Predictive Ability (EPA) is not rejected at a predefined confidence level — in our case, 85%. The EPA test is applied using an arbitrary loss function, which allows for the evaluation of models across different dimensions, such as punctual forecasts.This method effectively eliminates the least effective and redundant models from consideration. Among the available loss functions, our preferred choice is the mean squared error, which provides a robust measure for assessing the predictive accuracy of the models under review.

Figure 8: Model Confidence Set: Selected Models



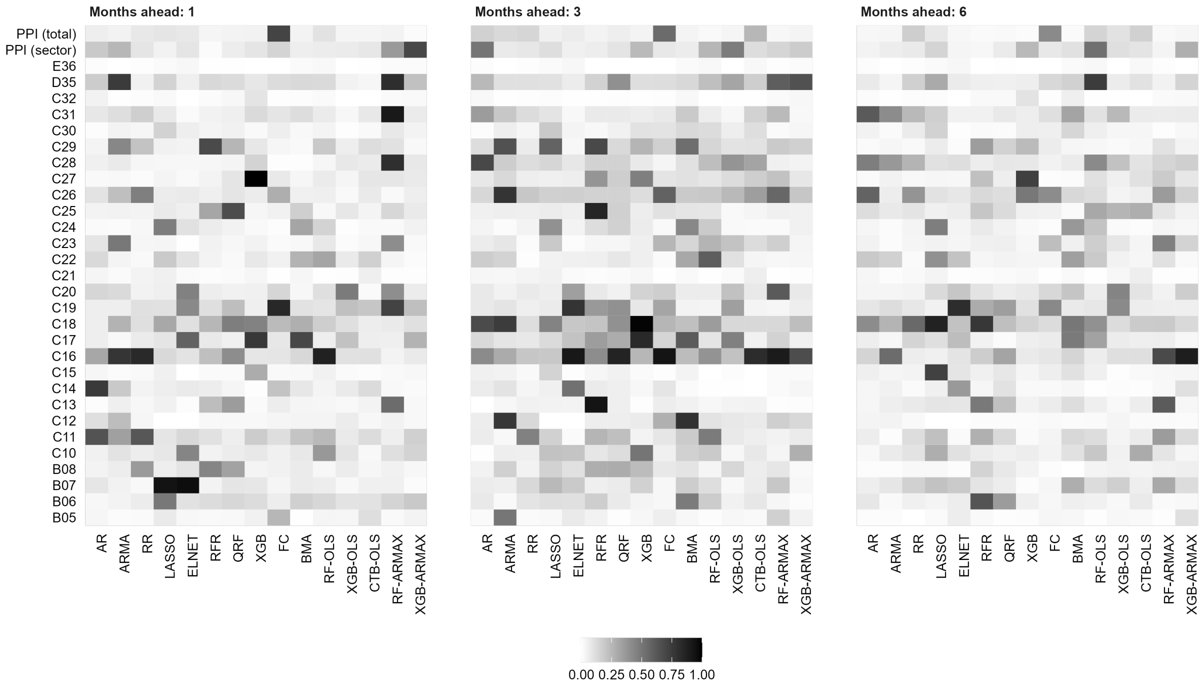
Note: Gray area stands for inclusion of the model in the superior model set

Figure 8 displays the outcomes of the MCS exercise, where the gray area represents models included in the MCS model set. This visual representation reveals that the approach deems most models to be robust in terms of their forecasting capabilities, as evidenced by the minimal number of blank cells. To synthesize the information from this model set effectively, we plan to compute a simple mean across the forecasts produced by the selected set of models.

The second approach builds upon the previously introduced capabilities of **Bayesian Model Averaging (BMA models)**, which is adept at identifying the most effective models. However, this iteration aims to determine an optimal set of models that spans different sectors and forecasting horizons, correlating the models' predictions with observed real values. Figure 9 illustrates the posterior inclusion probabilities for these models across both dimensions. The darker the color, the higher the probability that a model should be included in the model set, indicating its relevance and effectiveness. As observed, the model space appears more diverse, with certain models showing a strong preference across individual sectors. However, there are notable exceptions, such as sector C16, where several models display high PIP (indicated by darker shades). For these sectors, it may be more beneficial to utilize a combination of models rather than relying on a single best model.

The last three approaches for model selection are straightforward. In addition to employing a simple **Mean** and **Median** of forecasts across all models for a given sector, we also employ the **Sector Best**

Figure 9: Bayesian Model Averging: Selected Models



Note: Colour of the area stands for the posterior inclusion probability of the model

Model approach, which selects the model with the lowest RMSE for each sector. The models with the lowest RMSE can be identified directly from the row values in Table 2.

We continue our analysis by calculating the relative RMSE of various methods compared to the ARMA model. The results are presented in Table 3. All forecast combination methods prove to be viable alternatives to the baseline, exhibiting similar performance patterns. However, while these methods succeed in a number of sectors, their performance on longer horizons remains suboptimal for some C sectors. Overall, the value added by forecast combination is modest. Bayesian Model Averaging and the Model Confidence Set approaches demonstrate somewhat greater success than simple Mean or Median combinations. Selecting the best model for each sector emerges as a particularly effective strategy. This tailored approach ensures that the most suitable model is applied based on sector-specific dynamics, thereby optimizing forecast accuracy. One substantial drawback in practical use is that the sample of best models varies depending on the forecast horizon under consideration. Despite this variability, this method remains a valuable benchmark for comparison.

4.5 Sectoral Aggregation

Another research question we are investigating is whether a lower level of disaggregation challenges our outcomes. Our initial analyses, conducted at the NACE Rev. 2 two-digit level, have revealed significant noise, leading to relatively weak forecasting performance in some sectors, even over longer horizons. To address this issue, we are now examining the forecasting performance of our shortlisted models for PPI subgroups at a higher level of disaggregation. This exercise aims to determine if different granularity improves the accuracy and reliability of our forecasts by reducing the noise associated with broader category analyses.

Table 3: Relative RMSEs—Forecast Combinations and Model Selection

	Months ahead: 1					Months ahead: 3					Months ahead: 6				
Sector	Best model	BMA	MCS	Mean	Median	Best model	BMA	MCS	Mean	Median	Best model	BMA	MCS	Mean	Median
B05	0.94*	1.06	0.97	1.01	0.99	0.95*	1.03	0.97	1.01	0.99	0.98	1.18	1.06	1.06	1.03
B06	0.76**	0.78**	0.79***	0.79***	0.77**	0.71***	0.73***	0.74***	0.75***	0.74***	0.71***	0.77***	0.75***	0.76***	0.76***
B07	0.88*	0.9*	0.91	0.91	0.9	0.88	0.87**	0.88*	0.88*	0.88*	0.89*	0.9*	0.9*	0.91	0.89*
B08	0.88	0.92	0.92	0.93	0.95	1	1.03	1.05	1.05	1.06	0.92	0.95	0.95	0.95	0.96
C10	0.86	0.87*	0.89	0.89	0.92	0.65**	0.66***	0.69***	0.69***	0.69***	0.58**	0.61***	0.62***	0.63***	0.6***
C11	1	1.05	1.07	1.07	1.06	0.92*	0.97	0.96	0.96	0.98	0.81**	0.83***	0.86**	0.86**	0.89**
C12	1	1.07	1.13	1.12	1.05	1	1.03	1.01	1.08	1.02	1	1.18	1.14	1.16	1.05
C13	0.89**	0.92**	1	1	1.02	0.81***	0.83***	0.87***	0.87***	0.89**	0.59***	0.64***	0.71***	0.72***	0.74***
C14	1	1.14	1.15	1.17	1.16	1	1.06	1.12	1.13	1.1	0.96	1.01	1.03	1.05	1.02
C15	0.9**	0.95	1.03	1.03	0.98	0.95	0.98	1	1.02	0.97	0.91**	0.98	0.97	1.03	0.96
C16	1	1.05	1.05	1.05	1.07	0.55***	0.69***	0.72***	0.72***	0.72***	0.59***	0.64***	0.65***	0.65***	0.65***
C17	0.84**	0.9**	0.89**	0.91*	0.91*	0.61***	0.72***	0.67***	0.71***	0.7***	0.46***	0.51***	0.48***	0.56***	0.51***
C18	0.99	1.03	1.04	1.04	1.07	0.91*	1.01	1.05	1.05	1.07	0.91**	0.97	0.96	0.96	0.96
C19	0.42***	0.48***	0.44***	0.5***	0.45***	0.4***	0.44***	0.41***	0.47***	0.43***	0.4***	0.44***	0.41***	0.47***	0.42***
C20	0.68***	0.71***	0.71***	0.77***	0.75***	0.45***	0.48***	0.46***	0.54***	0.5***	0.42***	0.46***	0.43***	0.52***	0.48***
C21	0.91***	1.01	1	1.02	1.02	0.95**	1.03	1	1.02	1.02	0.94***	1.05	1.04	1.04	1.03
C22	0.93*	0.92**	0.93*	0.93*	0.93*	0.72***	0.76***	0.78***	0.78***	0.77***	0.6***	0.67***	0.65***	0.68***	0.67***
C23	1	0.98	1.03	1.04	1.04	0.85*	0.86***	0.88***	0.88***	0.89**	0.71**	0.72***	0.79***	0.79***	0.81***
C24	0.71***	0.73**	0.76**	0.76**	0.74**	0.49***	0.53***	0.54***	0.56***	0.53***	0.45***	0.48***	0.49***	0.52***	0.48***
C25	0.9	0.97	0.93	0.94	0.95	0.76*	0.82	0.77*	0.77*	0.78*	0.63***	0.68***	0.67***	0.68***	0.67***
C26	0.95	1.07	1	1.03	1.07	0.83***	0.88**	0.89**	0.91*	0.96	0.77***	0.81***	0.83***	0.83***	0.87***
C27	0.91	0.93	0.99	0.99	1.03	0.9	0.87**	0.88**	0.88**	0.91*	0.77***	0.76***	0.81***	0.81***	0.83***
C28	0.96	0.96	1.04	1.06	1.08	0.95	0.98	0.98	0.99	1.02	0.73**	0.85**	0.84**	0.84**	0.86**
C29	1	1.1	1.14	1.16	1.2	1	1.13	1.17	1.19	1.19	0.91	0.98	1.03	1.04	1.04
C30	1	0.99	0.99	1.01	0.98	1	1.03	1	1.02	1	0.98	0.97	0.96	0.98	0.96
C31	0.95	0.99	1.04	1.05	1.07	1	0.99	1	1.01	1.05	0.92	0.93*	0.93	0.94	0.97
C32	0.97	0.99	1.01	1.02	1.02	0.99	1.06	1.01	1.04	1.03	0.95	0.99	1.01	1.02	1.02
D35	0.64***	0.74***	0.73**	0.73**	0.73**	0.7**	0.73***	0.74***	0.74***	0.75**	0.74**	0.75**	0.76**	0.76**	0.77**
E36	0.96	0.97	0.97	0.98	0.98	0.99	1.01	1	1	1	0.99	1.02	1	1.01	1.01
PPI (sector)	0.61***	0.68***	0.69***	0.73***	0.75**	0.62***	0.69***	0.64***	0.68***	0.68***	0.57***	0.63***	0.63***	0.66***	0.65***
PPI (total)	0.75**	0.76***	0.76***	0.79***	0.78***	0.71***	0.72***	0.72***	0.74***	0.74***	0.67***	0.67***	0.68***	0.69***	0.69***

Note: Expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: BMA—Bayesian Model Averaging, MCS—Model Confidence Set.

We aim to employ the Main Industrial Grouping (MIG) approach, which offers an alternative statistical breakdown of industrial economic activities. MIGs serve as an intermediary level between the broader NACE sections (Sections B, C, D, and E) and the more detailed divisions and groups. The broader NACE sections are often too generalized for meaningful analysis, while the 34 divisions within these sections are too numerous and heterogeneous to adequately represent industry development over time.

There are five MIGs defined as follows: (i) Intermediate Goods; (ii) Capital goods; (iii) Consumer Durables; (iv) Consumer Non-durables; and (v) Energy. These groupings vary significantly in size, with consumer durables being notably smaller than the others. Unfortunately, NACE Rev. 2 two-digit level subgroups cannot be directly aggregated into appropriate MIG sectors due to partial overlaps across multiple MIGs. However, these discrepancies are relatively minor, so for simplicity, we allocate each NACE Rev. 2 two-digit subgroup to the MIG sector where it holds the most significance. [12] The weights of NACE Rev. 2 two-digit sectors are aggregated to determine the relative weights of the MIG sectors. In instances where data were not available for specific MIG level disaggregation—for example, selling price expectations from the European Commission survey—we aggregated the data using these relative weights to ensure consistency across levels of analysis. Table 4 illustrates our simplified methodology for assigning individual subsectors to their respective MIG groups.

[12] `https://ec.europa.eu/economy_finance/db_indicators/surveys/documents/2010/bcs_nace_2_classification_en.pdf`

Table 4: Allocation of NACE Rev. 2 Sectors to MIG

MIG	NACE Rev. 2 sectors									
MIG Capital Goods	C25	C26	C28	C29	C30					
MIG Consumer Durables	C31	C32								
MIG Energy	B05	B06	C19	D35	E36					
MIG Intermediate Goods	B07	B08	C13	C16	C17	C20	C22	C23	C24	C27
MIG Non-durable Consumer Goods	C10	C11	C12	C14	C15	C18	C21			

In alignment with our baseline setup, we conducted an out-of-sample forecasting exercise, the results of which are presented in Table 5. Consistent with earlier findings, certain models outperform others, although the differences are relatively modest. Notably, the superiority of these models increases over time. Ridge Regression, for instance, excels among the penalized regression methods, while Bayesian Model Averaging leads overall. Both Hybrid ARMAX models achieve particular success in forecasting the MIG Energy and Intermediate groups, which are closely tied to commodity price developments. Similarly, Hybrid OLS methods demonstrate some predictive power for these same MIG groups. Generally, as observed in our baseline analysis, models are less successful in forecasting MIG Consumer Durables. Moreover, the performance in forecasting MIG Capital Goods and MIG Non-durable Consumer Goods at short time horizons does not significantly surpass that of the simple ARMA model. The production processes in these sectors are often high-value added, and as producers positioned later in the production chains, they experience price pressures only after these have been transmitted through the pipeline. We will address this specific issue in the next section.

Table 5: Relative RMSEs—MIG Sectors

	AR-based		Penalized regression			Tree-based			Factor/BMA		Hybrid OLS			Hybrid ARMAX	
Sector	AR	ARMA	RR	LASSO	ELNET	RFR	QRF	XGB	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX
Months ahead: 1															
PPI (sector)	0.99	1	0.77**	0.78**	0.79**	0.91*	0.99	0.86**	0.77**	0.74***	0.76***	0.76***	0.79**	0.71***	0.7***
MIG Capital goods	1.06	1	1.17	1.1	1.1	1.07	1.21	1.14	1.1	1.06	1.06	1.13	1.21	1.03	0.98
MIG Consumer durables	1.03	1	1.16	1.12	1.14	1	1.1	1.01	1.13	1.09	1.08	1.05	1.25	1	1.06
MIG Energy	0.97	1	0.69***	0.73***	0.74**	0.87**	0.93	0.85**	0.7***	0.7***	0.71***	0.71***	0.72***	0.69***	0.69***
MIG Intermediate goods	0.98	1	0.93	0.81**	0.82**	1.05	1.08	0.95	0.9	0.77***	0.79***	0.74***	0.95	0.78**	0.77**
MIG Non-durable consumer goods	0.99	1	1.02	0.93	0.93	0.99	1.04	0.88	1.02	0.92	0.9	0.93	1.02	0.92	0.95
Months ahead: 3															
PPI (sector)	1.01	1	0.68***	0.7***	0.7***	0.84***	0.92	0.8***	0.69***	0.65***	0.67***	0.65***	0.71***	0.73***	0.64***
MIG Capital goods	1.01	1	1.07	1.02	1.03	1.08	1.23	1.1	0.97	0.96	0.96	0.95	1.15	1.21	1.19
MIG Consumer durables	1	1	1.12	1.1	1.11	0.99	1.09	1	1.1	1.07	1.04	1.05	1.17	1.14	1.2
MIG Energy	0.98	1	0.64***	0.68***	0.68***	0.83***	0.88**	0.78***	0.65***	0.64***	0.66***	0.63***	0.68***	0.78**	0.67***
MIG Intermediate goods	1.02	1	0.67***	0.59***	0.6***	0.82*	0.86	0.74***	0.65***	0.55***	0.56***	0.58***	0.68***	0.53***	0.53***
MIG Non-durable consumer goods	1	1	0.8**	0.73***	0.75**	0.82**	0.89	0.77***	0.79*	0.7**	0.7***	0.7**	0.82*	0.71**	0.62**
Months ahead: 6															
PPI (sector)	0.99	1	0.61***	0.64***	0.64***	0.78***	0.82***	0.73***	0.61***	0.58***	0.6***	0.61***	0.65***	0.63***	0.64***
MIG Capital goods	1.04	1	0.87**	0.82***	0.84***	0.95	1.05	0.89**	0.82**	0.8***	0.78***	0.75***	0.87*	1.01	0.96
MIG Consumer durables	1.02	1	1.09	1.07	1.08	0.98	1.06	0.96	1.04	1.01	1	1	1.17	1.26	1.22
MIG Energy	0.96	1	0.61***	0.66***	0.66***	0.81***	0.83***	0.77***	0.62***	0.62***	0.63***	0.64***	0.66***	0.72***	0.72***
MIG Intermediate goods	1.04	1	0.59***	0.49***	0.51***	0.74***	0.79***	0.67***	0.61***	0.47***	0.49***	0.5***	0.58***	0.49***	0.47***
MIG Non-durable consumer goods	1.03	1	0.71**	0.64***	0.64***	0.76***	0.79***	0.69***	0.69**	0.6**	0.61***	0.6**	0.64**	0.6**	0.71*

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: AR - Autoregressive Model, ARMA - Autoregressive Moving Average, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX

Finally, we explore whether the level of disaggregation affects the forecasting accuracy of the total PPI. A prevalent question in economic forecasting literature is whether insights gained from disaggregated-level forecasts can enhance predictions at a more aggregated level. Theoretically, forecasting at the most disaggregated level should retain all available information, preventing losses that might occur during aggregation. However, data at the bottom level can also be inherently noisy and pose greater challenges for modeling and forecasting.

Table 6: Relative RMSEs—Sectoral Aggregation vs. PPI (total)

	Baseline			MIG		
Model	Months ahead: 1	Months ahead: 3	Months ahead: 6	Months ahead: 1	Months ahead: 3	Months ahead: 6
AR	0.78**	0.8**	0.83**	0.77**	0.83*	0.85**
ARMA	0.76**	0.77***	0.75***	0.75**	0.79**	0.79**
RR	0.76**	0.75**	0.75**	0.75**	0.73**	0.71**
LASSO	0.75**	0.75**	0.75**	0.74**	0.74**	0.73**
ELNET	0.73*	0.73*	0.73*	0.76	0.75	0.73*
RFR	0.73**	0.73**	0.75**	0.78**	0.77**	0.77**
QRF	0.73**	0.73**	0.75**	0.81**	0.8**	0.79**
XGB	0.73***	0.72***	0.7**	0.76***	0.77**	0.73**
FC	0.74**	0.72**	0.71***	0.78*	0.76**	0.72**
BMA	0.72**	0.71**	0.72**	0.71**	0.71**	0.67**
RF-OLS	0.71**	0.7**	0.71**	0.72**	0.72**	0.71**
XGB-OLS	0.72***	0.65***	0.71**	0.74**	0.68***	0.71**
CTB-OLS	0.73***	0.73**	0.7***	0.75**	0.75**	0.72**
RF-ARMAX	0.59**	0.69**	0.61***	0.68*	0.63**	0.63***
XGB-ARMAX	0.58**	0.54***	0.58***	0.66**	0.57***	0.68**
Best model (comb)	0.62*	0.67*	0.65**	0.7*	0.71*	0.69**
BMA (comb)	0.69**	0.74**	0.71***	0.71**	0.74**	0.71**
MCS (comb)	0.69**	0.69***	0.7***	0.71**	0.7**	0.69**
Mean (comb)	0.71**	0.71**	0.72**	0.73**	0.73**	0.73**
Median (comb)	0.73**	0.71**	0.71**	0.74**	0.72**	0.72**

Note: Expressed as a share of the RMSE for the PPI (total). ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: AR - Autoregressive Model, ARMA - Autoregressive Moving Average, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX, MCS - Model Confidence Set.

The disaggregated approach, referred to as PPI (sector) in our baseline analysis, demonstrated superior performance compared to the simple ARMA model across various model setups. This encouraging outcome suggests that sector-based forecasting can significantly enhance predictive accuracy. More formally, we assessed the performance of sector-based forecasts of PPI against aggregate forecasts. The results, presented in Table 6, clearly favor our sectoral approach, showing notable statistical significance. We differentiate between two levels of disaggregation — NACE Rev. 2 (Baseline) and Main Industrial Groupings (MIG) — across all models. The sectoral approach consistently outperforms in forecasting total PPI, with gains in terms of RMSE reaching as much as nearly 40%. These improvements are particularly pronounced for Hybrid ARMAX models, which excel in forecasting energy subsectors. The positive effect can be then attributed to the weight of Energy in the total PPI, as energy subsectors together contribute nearly one-third to the total index. Even when employing a factor model, which has been identified as the most effective framework for modeling total PPI, the largest gains are observed with Hybrid OLS and Hybrid ARMAX models. Additionally, model combinations have proven to be quite beneficial. This analysis underscores the robustness and utility of a sectoral forecasting approach, particularly when detailed and sector-specific data are utilized effectively.

4.6 Accounting for Pipeline Pressures

Individual sectors represented in the Producer Price Index reflect prices at various stages of the production chain, each capturing the dynamic nature of pricing from raw materials to finished products. For instance, the crude materials encompasses prices at the earliest stage of production, such as crude petroleum. Progressing along the production chain, the Intermediate goods includes prices for products like synthetic rubber, which is synthesized from crude petroleum. Subsequently, the Finished goods comprises prices of end products like tires, manufactured from rubber. This nuanced distinction among sectors, while similar to the Main Industrial Grouping approach discussed earlier, provides a deeper layer of analysis by directly reflecting the progression within the production chain. Unlike the broader categories outlined in the MIG framework, this detailed sectoral breakdown facilitates a more granular examination of how price shifts in one stage of production can ripple through to subsequent stages, offering a more precise understanding of the economic interdependencies across different segments of industry.

The progression from raw materials to finished products illustrates a cascade of sectoral shocks, which we term "Pipeline Pressures". These pressures impact the pricing dynamics at each stage of production. At final stage they influnce the consumer inflation rate. For example, changes in the prices of car tires, a key component in vehicle manufacturing, affect the costs of vehicle transportation services and the cost of vehicles paid by consumers (i.e. both services and goods component of CPI). By examining this cascade effect, we can assess how shocks in one sector propagate through the production chain, affecting prices in downstream sectors.

The presence and impact of pipeline pressures has been widely discussed in recent analytical work by Czech National Bank (GEV-2022) and European Central Bank (ECB-2021, ECB-2023). More theoretical background with relevant literature in this field can be found in Smets et al. (2019), which provides an evidence that producer prices of downstream sectors (e.g. Manufacturing) are strongly subject to price developments in upstream sectors (e.g. Mining). Moreover, this study shows that some pipeline pressures manifest themselves quickly whereas others take time to build, manily due to heterogeneous price stickiness of sectors. This effects explain a large share of the heterogeneity observed in the persistence of disaggregate inflation data. Moreover, pipeline pressures are documented to be a key source of disaggregate/headline inflation volatility.

To account for pipeline pressures, our latest forecasting exercise incorporates lagged developments from other sectors into each NACE Rev. 2 sector analysis. We consider a maximum of two lags, intentionally avoiding contemporaneous effects from other sectors to minimize confounding influences. The results, which are compared relative to the ARMA model as in our baseline setup, are presented for one-month-ahead forecasts in Table 7. Results for three and six-month-ahead forecasts can be found in Tables C.3 and C.4 in Appendix C. As anticipated, the forecasting accuracy for Finished Goods (approximately sectors C25 to C32) improved significantly. Machine learning methods now outperform our benchmark even on the shortest horizon, with many differences being statistically significant. This improvement underscores the necessity of accounting for pipeline pressures in Final Goods sectors, where pricing policies reflect changes in input prices with a time lag and often react only to price alterations in earlier production stages. While improvements are also observed in several intermediate sectors (sectors C11 to C18, C22, and C23), they are not as marked. Conversely, there is a noticeable deterioration in the forecasting performance of Energy sectors (e.g., C19, D35) and other sectors supplying crude materials (e.g., B05 to B07).

Table 7: Relative RMSEs for One-month-ahead Forecasts—Pipeline Pressures

	AR-based		Penalized regression			Tree-based			Factor/BMA		Hybrid OLS			Hybrid ARMAX	
Sector	AR	ARMA	RR	LASSO	ELNET	RFR	QRF	XGB	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX
PPI (total)	0.97	1	0.78**	0.79**	0.79**	0.88**	0.97	0.85**	0.75**	40.26	0.77***	0.81**	0.9	0.85*	0.9
PPI (sector)	0.99	1	0.78***	0.79**	0.78**	0.85**	0.9	0.84*	0.81**	21.57	0.73***	0.76**	0.9	0.78**	0.76**
B05	0.96	1	0.97	1.01	0.98	1.02	0.97	1.08	1.19	6.78	1.03	1.19	1.06	1.22	1.22
B06	0.97	1	0.87**	0.79***	0.8***	0.85***	0.86***	0.91	0.93	1.32	0.78**	0.84**	0.94	0.82*	0.79**
B07	1.01	1	1.06	0.9*	0.9*	0.98	0.93	0.95	1.42	3.46	0.95	1.16	1.32	0.98	1.06
B08	1.06	1	0.92	0.93	0.91	0.9	0.92	0.96	0.96	312.84	0.9	0.84*	1.02	0.96	1.04
C10	1	1	0.88	0.91*	0.93	1.01	1.03	1	0.98	84.97	0.86**	0.91	1.05	0.99	0.94
C11	1	1	1.02	0.99	1	0.97	1.05	1.01	1.27	45.48	1	1.06	1.19	0.99	1.15
C12	1.11	1	1.07	1	1	1.56	1.04	2.59	3.23	28.86	1.22	1.95	1.7	1.78	1.54
C13	1.04	1	1.04	1.11	1.11	0.95	1.01	1.1	1.09	42.58	0.99	1.05	1.22	1.07	1.16
C14	1.08	1	1.1	1.15	1.16	0.97	1.02	1.15	1.55	50.41	1.04	1.26	1.19	1.09	1.16
C15	1.02	1	0.94*	0.97	1	0.91**	0.91***	0.9**	1.45	62.38	0.93	1.15	1.1	1.11	1.21
C16	1.1	1	1.18	1.1	1.11	1.2	1.16	1.34	1.27	17.83	1.15	1.21	1.32	1.24	1.37
C17	1	1	0.96	0.94	0.93	1.18	1.16	1.09	1.02	65.8	0.93	0.95	1.18	1	1.04
C18	1.05	1	0.99	1.07	1.06	0.92*	1.04	0.96	1.01	24.19	0.95	1.04	1.1	1	0.94
C19	0.97*	1	0.57***	0.48***	0.48***	0.66***	0.64***	0.62***	0.78***	2.08	0.45***	0.43***	0.93	0.43***	0.44***
C20	0.98*	1	0.85*	0.78***	0.8***	0.97	1.02	0.99	0.92	11.21	0.76***	0.75***	0.86*	0.76***	0.7***
C21	0.91***	1	1.02	1.06	1.06	1.04	0.95**	1.02	1.7	61.46	1.03	1.27	1.39	1.25	1.31
C22	0.97**	1	0.92*	0.95	0.96	0.92*	0.97	0.85	1.12	47.16	0.89**	0.95	1.07	0.97	0.91
C23	1.05	1	1.02	1.07	1.06	1.07	1.09	1.09	1.02	32.26	1	0.99	1.12	1.1	1.04
C24	0.97	1	0.88	0.74**	0.75***	0.98	1	0.9	1.11	10.46	0.75**	0.8*	1.15	0.75**	0.8*
C25	0.97	1	0.98	1.1	1.08	1.02	1.06	0.95	1.15	694.91	1	1.19	1.12	1.13	1.08
C26	0.95	1	1.38	1.27	1.31	1.22	1.1	1.18	1.48	52.03	1.26	1.52	1.28	1.4	1.24
C27	1.04	1	1	1.04	1.05	0.93**	0.98	0.89*	1.19	132.52	0.97	1.09	1.28	0.98	1.13
C28	1.05	1	1.03	1.04	1.05	0.96*	1	1.02	1.21	91.26	1.01	1.13	1.31	1.12	1.08
C29	1.17	1	1.14	1.15	1.17	0.98	1.07	1.14	1.32	135.78	1.07	1.22	1.14	1.16	1.23
C30	1.08	1	0.94	0.91**	0.93*	0.92	0.92**	1.2	1.67	64.89	1.07	1.28	1.41	1.53	1.46
C31	1.07	1	1.03	1.06	1.06	0.99	1.04	0.96	1.2	356.21	1	1.08	1.19	1.17	1.08
C32	1.02	1	1.05	1.08	1.03	0.97	0.94*	1.07	1.37	233.68	1.02	1.21	1.19	1.16	1.24
D35	0.93	1	0.79**	0.76*	0.75**	0.82**	0.85*	0.84*	0.83	2.37	0.74**	0.77*	0.93	0.81*	0.81*
E36	1.11	1	0.97	0.98	0.98	0.99	0.97	1.06	1.18	967.22	0.98	1.03	1.14	1.04	1

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: AR - Autoregressive Model, ARMA - Autoregressive Moving Average, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX

Focusing on the comparison across methods, those based on Random Forest (RFR, QRF, RF-OLS) distinctly outperform other approaches. Penalized regression methods also demonstrate their effectiveness by aptly selecting relevant drivers for each sector. While Hybrid ARMAX methods continue to achieve success in some sectors, their overall performance has declined. The results from Bayesian Model Averaging (BMA) have significantly deteriorated, likely due to the increased complexity of the model setup while retaining the default configuration. To mitigate overfitting in this context, a different approach would be necessary, involving adjustments in both the selected priors and the variables considered. Additionally, the factor model, although losing some predictive strength compared to our baseline, remains a pertinent forecasting tool for total PPI. These observations are consistent across longer forecasting horizons, where most models substantially outperform our ARMA benchmark.

5. Conclusion

This study has tried to address several gaps in the empirical literature concerning the forecasting of disaggregated producer prices, particularly through the application of machine learning techniques alongside traditional econometric models. Our findings indicate that no single forecasting model consistently outperforms others across all sectors, underscoring the complexity and diversity of pricing dynamics within the euro area. This necessitates a tailored approach, leveraging the strengths of various forecasting methods to capture the unique characteristics of each sector. Aside Bayesian Model Averaging machine learning methods like Random Forest or XGBoost have proven particularly effective, especially when combined with traditional approaches in our hybrid models. Notably, the Hybrid RF-ARMAX and XGB-ARMAX models, introduced as innovative solutions in this study, emerge as robust options for forecasting short-term PPI inflation across various sectors, particularly those influenced by commodity prices.

The integration of machine learning methods has not only improved forecasting accuracy but has also offered deeper insights into the dynamic interplay of factors that influence producer prices. Our forecasting exercise has highlighted diverse pricing strategies linked to commodity prices, autoregressive behavior, or a mixture of both, with pipeline pressures being especially pertinent to final goods. Employing a mixture of a wide range of models has proven to be a successful strategy in managing the varied pricing behavior at the sectoral level.

However, the study also underscores the challenges associated with modeling at a highly disaggregated level, primarily due to the inherent noise in such detailed data. While a disaggregated approach provides nuanced insights into sector-specific shocks and their propagation through the production chain, it complicates the forecasting process. Ultimately, our analysis indicates that a lower level of disaggregation does not necessarily lead to a substantial loss in forecasting accuracy.

Nevertheless, despite these challenges, the results highlight that a disaggregated approach consistently delivers superior forecasting performance, particularly in capturing unique sectoral dynamics that aggregated models may overlook. This superiority is particularly notable for short-term forecasting and sectors with high sensitivity to commodity and energy price fluctuations. Our findings align with the broader literature, which also supports the effectiveness of disaggregated approach to forecasting inflation like Bermingham and D'Agostino (2014) or Ibarra (2012).

In conclusion, machine learning methods have emerged as a compelling alternative to traditional econometric approaches in our analysis. However, it is essential not to forsake established methodologies for newer ones without careful consideration. Instead, we should aim to maintain a diversified portfolio of methods, leveraging the distinct advantages of each based on the specific characteristics of the dataset. Machine learning methods have demonstrated their strength in managing large datasets with potentially correlated variables. Yet, when combined with traditional approaches, they also adeptly address time series with strong temporal dependencies, such as inflation data.

References

ALMOSOVA, A. AND N. ANDRESEN (2023): "Nonlinear inflation forecasting with recurrent neural networks." *Journal of Forecasting,* 42(2):240–259.

ARAUJO, G. S. AND W. P. GAGLIANONE (2022): *Machine learning methods for inflation forecasting in Brazil: new contenders versus classical models.* Banco Central do Brasil.

AUER, R. A., A. A. LEVCHENKO, AND P. SAURÉ (2019): "International inflation spillovers through input linkages." *Review of Economics and Statistics,* 101(3):507–521.

BAI, J. AND S. NG (2002): "Determining the number of factors in approximate factor models." *Econometrica,* 70(1):191–221.

BARKAN, O., J. BENCHIMOL, I. CASPI, E. COHEN, A. HAMMER, AND N. KOENIGSTEIN (2023): "Forecasting CPI Inflation Components with Hierarchical Recurrent Neural Networks." *International Journal of Forecasting.*

BENECKÁ, S. AND L. KOMAREK (2018): "International reserves: Facing model uncertainty." *Economic Systems,* 42(3):523–531.

BERK, R. A. (2008): *Statistical learning from a regression perspective.* Springer.

BERMINGHAM, C. AND A. D'AGOSTINO (2014): "Understanding and forecasting aggregate and disaggregate price dynamics." *Empirical Economics,* 46(2):765–788.

BHATTACHARYA, P. S. AND D. D. THOMAKOS (2008): "Forecasting industry-level CPI and PPI inflation: Does exchange rate pass-through matter?" *International Journal of Forecasting,* 24(1):134–150.

BOARETTO, G. AND M. C. MEDEIROS (2023): "Forecasting inflation using disaggregates and machine learning." *arXiv preprint arXiv:2308.11173.*

BRÁZDIK, F., T. HLÉDIK, Z. HUMPLOVÁ, I. MARTONOSI, K. MUSIL, J. RYŠÁNEK, T. ŠESTOŘÁD, J. TONNER, S. TVRZ, AND J. ŽÁČEK (2020): *The g3+ Model: An Upgrade of the Czech National Bank's Core Forecasting Framework.* Czech National Bank, Economic Research Department.

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

CHAKRABORTY, C. AND A. JOSEPH (2017): "Machine learning at central banks." *BoE Staff Working Paper No. 674, Bank of England.*

CHOUDHARY, M. A. AND A. HAIDER (2012): "Neural network models for inflation forecasting: an appraisal." *Applied Economics,* 44(20):2631–2635.

CORNILLE, D. AND M. DOSSCHE (2006): "The patterns and determinants of price setting in the Belgian industry." *Working Paper 82, National Bank of Belgium.*

DE SANTIS, R. A. (2024): "Supply Chain Disruption and Energy Supply Shocks: Impact on Euro-Area Output and Prices." *International Journal of Central Banking,* 20(2).

DIEBOLD, F. X. AND R. S. MARIANO (2002): "Comparing predictive accuracy." *Journal of Business & Economic Statistics,* 20(1):134–144.

ECB-2021: "Economic Bulletin 05/2021." Available from: (https://www.ecb.europa.eu/pub/economic-bulletin/html/eb202105.en.html)

ECB-2023: "Economic Bulletin 03/2023." Available from: (https://www.ecb.europa.eu/press/economic-bulletin/focus/2023/html/ecb.ebbox202303_05~e6d2439ece.en.html)

EUROSTAT (2012): "Handbook on Industrial Producer Price Indices (PPI)."

FRIEDMAN, J. H. (2001): "Greedy function approximation: a gradient boosting machine." *Annals of Statistics,* 1189–1232.

GARCIA, M. G., M. C. MEDEIROS, AND G. F. VASCONCELOS (2017): "Real-time inflation forecasting with high-dimensional models: The case of Brazil." *International Journal of Forecasting,* 33(3):679–693.

GARETH, J., W. DANIELA, H. TREVOR, AND T. ROBERT (2013): *An introduction to statistical learning: with applications in R.* Springer.

GAUTIER, E. (2008): "The behaviour of producer prices: evidence from French PPI micro data." *Empirical Economics,* 35:301–332.

GEV-2022: "Global Economic Outlook 09/2022." Available from: (https://www.cnb.cz/export/sites/cnb/en/monetary-policy/.galleries/geo/geo_2022/gev_2022_09_en.pdf)

GUNN, S. R. ET AL. (1998): "Support vector machines for classification and regression." *ISIS Technical Report,* 14(1):5–16.

HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): "The model confidence set." *Econometrica,* 79(2):453–497.

HASTIE, T., R. TIBSHIRANI, J. H. FRIEDMAN, AND J. H. FRIEDMAN (2009): *The elements of statistical learning: data mining, inference, and prediction.* Springer.

HOERL, R. W. (2020): "Ridge regression: A historical context." *Technometrics,* 62(4):420–425.

HYNDMAN, R. J. AND G. ATHANASOPOULOS (2018): *Forecasting: principles and practice.* OTexts.

IBARRA, R. (2012): "Do disaggregated CPI data improve the accuracy of inflation forecasts?" *Economic Modelling,* 29(4):1305–1313.

JACOBSON, T. AND S. KARLSSON (2004): "Finding good predictors for inflation: A Bayesian model averaging approach." *Journal of Forecasting,* 23(7):479–496.

JAMES, G., D. WITTEN, T. HASTIE, R. TIBSHIRANI, ET AL. (2013): *An introduction to statistical learning.* Springer.

JIANG, J. AND D. KIM (2013): "Exchange rate pass-through to inflation in China." *Economic Modelling,* 33:900–912.

KLEINBERG, J., J. LUDWIG, S. MULLAINATHAN, AND Z. OBERMEYER (2015): "Prediction policy problems." *American Economic Review,* 105(5):491–495.

KOHLSCHEEN, E. (2021): "What does machine learning say about the drivers of inflation?" *BIS Working Papers 980, Bank for International Settlements.*

KOOP, G. AND D. KOROBILIS (2012): "Forecasting inflation using dynamic model averaging." *International Economic Review,* 53(3):867–886.

LABELLE, J. AND A. M. SANTACREU (2022): "Global supply chain disruptions and inflation during the COVID-19 pandemic." *Review, Federal Reserve Bank of St. Louis.*

LENZA, M., I. MOUTACHAKER, AND J. PAREDES (2023): "Density forecasts of inflation: A quantile regression forest approach." *ECB Working Paper Series 2830, European Central Bank.*

MCADAM, P. AND P. MCNELIS (2005): "Forecasting inflation with thick models and neural networks." *Economic Modelling,* 22(5):848–867.

MEDEIROS, M. C. AND E. F. MENDES (2016): "L1-regularization of high-dimensional time-series models with non-Gaussian and heteroskedastic errors." *Journal of Econometrics,* 191 (1):255–271.

MEDEIROS, M. C., G. F. VASCONCELOS, Á. VEIGA, AND E. ZILBERMAN (2021): "Forecasting inflation in a data-rich environment: the benefits of machine learning methods." *Journal of Business & Economic Statistics,* 39(1):98–119.

MEINSHAUSEN, N. AND G. RIDGEWAY (2006): "Quantile regression forests." *Journal of Machine Learning Research,* 7(6).

MOSHIRI, S. AND N. CAMERON (2000): "Neural network versus econometric models in forecasting inflation." *Journal of Forecasting,* 19(3):201–217.

MYERS, R. J., S. R. JOHNSON, M. HELMAR, AND H. BAUMES (2018): "Long-run and short-run relationships between oil prices, producer prices, and consumer prices: What can we learn from a permanent-transitory decomposition?" *Quarterly Review of Economics and Finance,* 67:175–190.

NAKAMURA, E. (2005): "Inflation forecasting using a neural network." *Economics Letters,* 86(3): 373–378.

PARANHOS, L. (2021): "Predicting Inflation with Neural Networks." *Working Paper 1344, University of Warwick, Department of Economics.*

SCHNORRENBERGER, R. AND G. V. MOURA (2024): "Harnessing Machine Learning for Real-Time Inflation Nowcasting." In *Annual Conference,* number 1/12.

SIDAOUI, J., C. CAPISTRÁN, D. CHIQUIAR, AND M. RAMOS-FRANCIA (2010): "On the predictive content of the PPI on CPI inflation: the case of Mexico." *BIS Papers No 49 , Bank for International Settlements.*

SMETS, F., J. TIELENS, AND J. VAN HOVE (2019): "Pipeline Pressures and Sectoral Inflation Dynamics." *Working Paper 351, National Bank of Belgium.*

STOCK, J. H. AND M. W. WATSON (2002): "Forecasting using principal components from a large number of predictors." *Journal of the American Statistical Association,* 97(460):1167–1179.

THEOHARIDIS, A. F., D. A. GUILLÉN, AND H. LOPES (2023): "Deep learning models for inflation forecasting." *Applied Stochastic Models in Business and Industry.*

TIBSHIRANI, R. (1996): "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B (Methodological),* 58(1):267–288.

WEI, S.-J. AND Y. XIE (2018): "The Wedge of the Century: Understanding a Divergence between CPI and PPI Inflation Measures." *Working Paper 24319, National Bureau of Economic Research.*

WRIGHT, J. H. (2009): "Forecasting US inflation by Bayesian model averaging." *Journal of Forecasting,* 28(2):131–144.

YE, Y.-F., H. CAO, L. BAI, Z. WANG, AND Y.-H. SHAO (2013): "Exploring Determinants of Inflation in China based on L1-E-Twin Support Vector Regression." *Procedia Computer Science,* 17:514–522.

ZHAO, L. L., B. WANG, J. MBACHU, AND T. EGBELAKIN (2020): "Using artificial neural networks to forecast producer price index for New Zealand." *International Journal of Internet Manufacturing and Services,* 7(3):191–215.

ZOU, H. (2006): "The adaptive lasso and its oracle properties." *Journal of the American Statistical Association,* 101(476):1418–1429.

ZOU, H. AND T. HASTIE (2005): "Regularization and variable selection via the elastic net." *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 67(2):301–320.

Appendix A: Data

The Producer Price Index (PPI) is a measure of price changes from the perspective of the seller and is computed as a weighted average of product prices. The PPI is calculated for three distinct market segments: total, domestic, and non-domestic, with separate indices compiled based on the destination of the products sold. According to the guidelines, the appropriate price for PPI calculation is the basic price, which excludes VAT, similar deductible taxes, duties, and taxes on invoiced goods and services. Additionally, any subsidies received by the producer should be included in the basic price. It is important that the prices collected reflect orders booked at the time of the order placement rather than at the time the commodities are dispatched from the factory gates (Eurostat, 2012). Eurostat collects PPI data at the 2-digit level of the NACE classification from national statistical offices. The original monthly data are indexed to 2015 = 100 and are not adjusted for calendar or seasonal variations. Therefore, before any estimation, the individual indices are seasonally adjusted using the X-13 Tramo/Seats method to ensure accuracy and consistency in analysis.

Table A.1: Description of Variables

No.	Code	Variable	Description	Seasonal adj.	Source	Unit	Transformation
1	ppi	Producer price index		NSA	Eurostat	index, 2015=100	log diff
2	ip	Industrial production		SCA	Eurostat	index, 2015=100	log diff
3	esi_orders	Industrial confidence	Assessment of order-book levels	SA	EC	balance	diff
4	esi_stocks	Industrial confidence	Assessment of stocks of finished products	SA	EC	balance	diff
5	esi_production	Industrial confidence	Production expectations for months ahead	SA	EC	balance	diff
6	esi_selprice	Industrial confidence	Selling price expectations for months ahead	SA	EC	balance	diff
7	flp_mat	Factors limiting production	Material and/or equipment	SA	EC	balance	diff
8	flp_lab	Factors limiting production	Labor	SA	EC	balance	diff
9	flp_dem	Factors limiting production	Demand	SA	EC	balance	diff
10	Food	Commodity index	Food	NSA	Bloomberg	index, 2005=100	log diff
11	Metals	Commodity index	Metals	NSA	Bloomberg	index, 2005=100	log diff
12	Brent	Oil price	Brent	NSA	Bloomberg	USD	log diff
13	NaturalGasWB	Natural gas price		NSA	Bloomberg	USD	log diff
14	ElectricityDE	Electricity price		NSA	Bloomberg	USD	log diff
15	Sugar	Sugar price		NSA	Bloomberg	USD	log diff
16	LiveCattle	Live cattle price		NSA	Bloomberg	USD	log diff
17	Cotton	Cotton price		NSA	Bloomberg	USD	log diff
18	Cocoa	Cocoa price		NSA	Bloomberg	USD	log diff
19	OrangeJuice	Orange juice price		NSA	Bloomberg	USD	log diff
20	Rice	Rice price		NSA	Bloomberg	USD	log diff
21	BalticDryIndex	Baltic Dry Index		NSA	Bloomberg	USD	log diff
22	GSCPI	Global Supply Chain Pressure Index		NSA	New York Fed	stdev from avg	diff
23	PMI	Purchasing Managers' Index	Manufacturing	SA	S&P Global	diffusion index	diff
24	CR	New car registrations		NSA	Bloomberg	unit	log diff

Note: The transformation column indicates the transformation necessary to ensure stationarity. Seasonal adjustment and unit relate to the original series.

Table A.2: PPI Sectors: Description and Summary Table

NACE Rev.2 Code	Description	Adjusted weight	Mean	SD	Max	Min
D35	Electricity, gas, steam and air conditioning supply	20.04	0.41	2.59	16.66	-14.42
C10	Manufacture of food products	13.45	0.20	0.55	4.13	-2.70
C25	Manufacture of fabricated metal products, except machinery and equipment	6.95	0.17	0.36	2.23	-0.51
C19	Manufacture of coke and refined petroleum products	6.77	0.35	5.13	21.62	-21.30
C29	Manufacture of motor vehicles, trailers and semi-trailers	6.20	0.10	0.17	1.21	-0.23
C20	Manufacture of chemicals and chemical products	5.55	0.22	1.07	4.77	-3.65
C28	Manufacture of machinery and equipment n.e.c.	5.20	0.16	0.21	1.69	-0.29
C24	Manufacture of basic metals	4.35	0.28	1.45	5.47	-5.58
C22	Manufacture of rubber and plastic products	3.53	0.13	0.36	2.21	-1.17
C23	Manufacture of other non-metallic mineral products	3.33	0.20	0.39	3.22	-1.41
C27	Manufacture of electrical equipment	3.09	0.13	0.27	1.75	-0.46
C26	Manufacture of computer, electronic and optical products	2.34	-0.19	0.42	1.04	-2.62
C17	Manufacture of paper and paper products	2.17	0.15	0.71	4.01	-1.73
C11	Manufacture of beverages	2.04	0.17	0.26	2.11	-0.49
C30	Manufacture of other transport equipment	2.00	0.10	0.21	1.50	-0.96
C16	Manufacture of wood and of products of wood and cork	1.68	0.17	0.65	6.00	-0.99
C18	Printing and reproduction of recorded media	1.67	0.05	0.37	3.34	-0.67
C21	Manufacture of basic pharmaceutical products and preparations	1.64	-0.01	0.24	0.94	-1.53
C31	Manufacture of furniture	1.45	0.18	0.25	2.13	-0.18
E36	Water collection, treatment and supply	1.16	0.18	0.24	2.54	-0.84
C14	Manufacture of wearing apparel	1.12	0.08	0.23	1.45	-1.08
C32	Other manufacturing	1.03	0.13	0.25	1.85	-0.78
C13	Manufacture of textiles	0.98	0.13	0.29	2.22	-0.35
B06	Extraction of crude petroleum and natural gas	0.62	0.48	6.66	26.06	-42.26
B08	Other mining and quarrying	0.57	0.22	0.39	4.16	-1.15
C15	Manufacture of leather and related products	0.57	0.16	0.35	2.12	-0.92
C12	Manufacture of tobacco products	0.33	0.30	0.95	8.54	-1.02
B05	Mining of coal and lignite	0.14	0.35	2.26	15.79	-8.77
B07	Mining of metal ores	0.02	0.30	3.76	15.78	-16.68
B-E36	Industry	100.00	0.23	0.93	5.28	-3.36

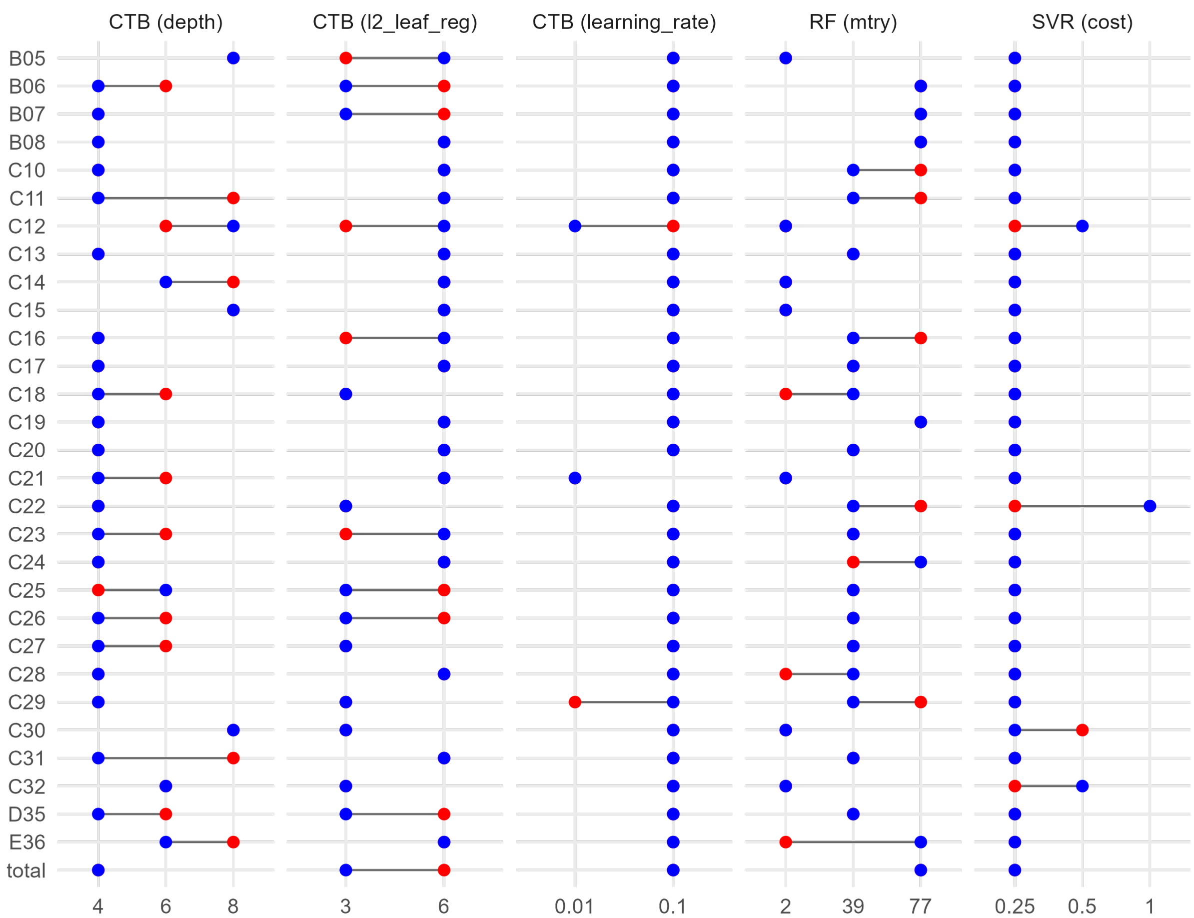
Note: Weights adjusted for missing data on C33. For visibility, statistics computed on 100 * log differences

Appendix B: Parameter Grid for Machine Learning Methods

One of the fundamental challenges in applying machine learning techniques is the optimal tuning of model parameters. Fine-tuning parameters is crucial for data analysts utilizing advanced machine learning methods, as it significantly impacts model performance. In this section, we explore how the choice of parameters varies across different times and sectors, specifically examining whether there is a notable difference in model parameters for each sector and how these preferences have shifted in the post-COVID period compared to the overall period under review. While some methods, like penalized regression, inherently incorporate parameter optimization, others, such as boosting tree models, random forests, and support vector machines, require a more deliberate approach through cross-validation, respecting time series nature of data.

We illustrate the preferred choices of key parameters for a set of three models (CTB, RF and SVR) in Figure B.1. In the figure, blue points represent parameter selections pre-COVID, while red points indicate values for the entire sample within a fixed window. This distinction is intended to reflect the practical considerations forecasters face when predicting parameter evolution in periods of high uncertainty. It also acknowledges the challenges and costs associated with continuously adapting parameters to changes within individual sectors over time.

Figure B.1: Model Parameters Based on Selection of the Optimal Model (CTB, RF, SVR)



Note: Method with the name of the parameter in brackets, blue points show the results for pre-pandemic period and red stand for values for the entire sample.

As for CatBoost, we consider the following parameters:

- depth: maximum depth of a tree
- l2_leaf_reg: L2 regularization coefficient used for leaf value calculation
- learning_rate: the learning rate used for reducing the gradient step

Figure B.1 suggests that a lower tree depth, a higher L2 regularization coefficient, and an increased learning rate are the preferred settings for CatBoost, regardless of the specific sectors or time periods analyzed. These parameter choices indicate a trend towards models that prevent overfitting through more significant regularization and faster convergence rates, enhancing generalization across diverse datasets.

In the case of Random Forest, the mtry parameter represents the number of variables randomly sampled as candidates at each split. For Support Vector Regression, the cost parameter determines the penalty for constraint violations. For Random Forest, the preferred mtry value is often 39, close to the default setup, and typically remains unchanged even when considering the entire period. For Support Vector Regression, grid search results for the cost parameter suggest that the default value should be adjusted from one to 0.25, a preference that persists throughout the analysis period. This adjustment in the cost parameter has already been incorporated into our baseline.

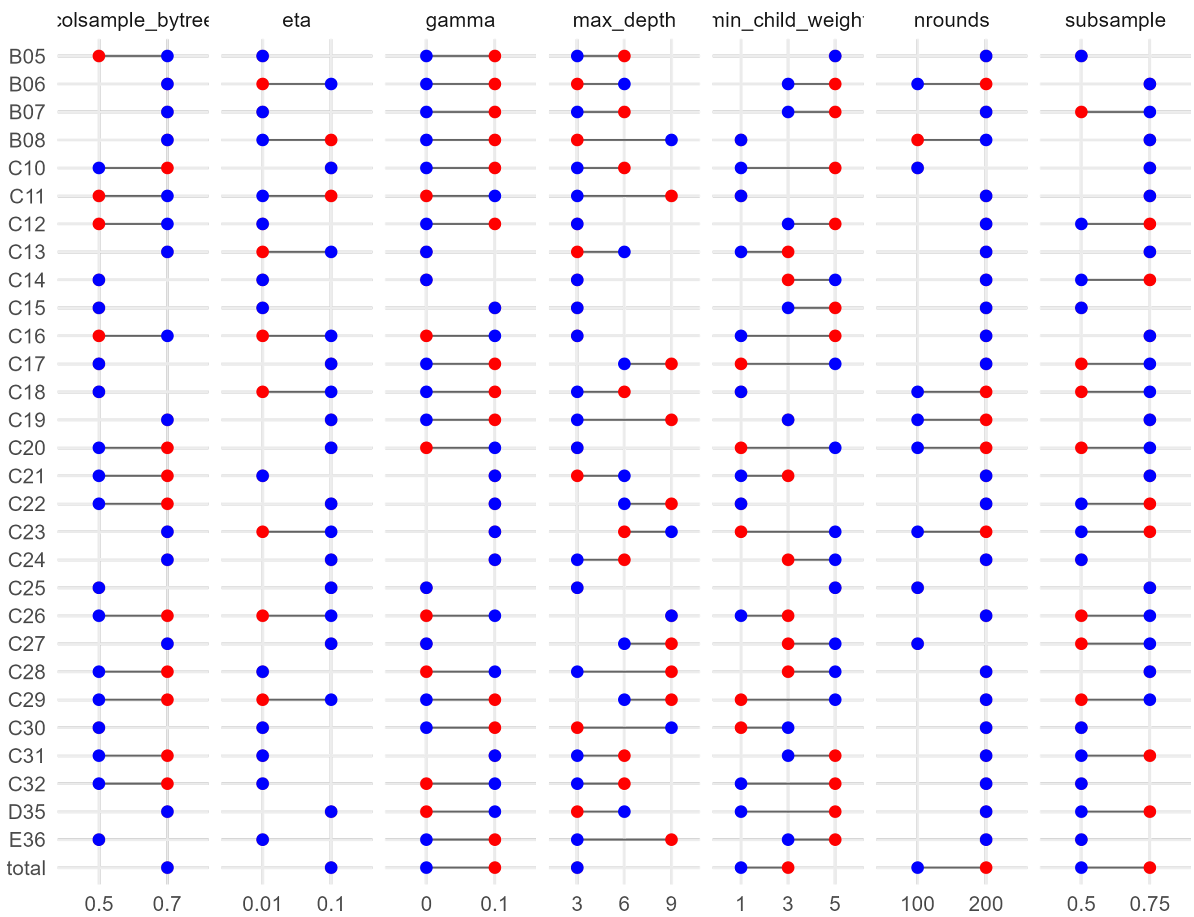
To summarise, the analysis of these parameters across various sectors and models yields promising results. Despite some deviations in parameter preferences across sectors, the overall stability in parameter performance indicates robustness and adaptability of these models to handle sector-specific dynamics effectively.

Moving to XGBOOST, the parameters we have investigated are defined as follows:

- colsample_by_tree: subsample ratio of columns when constructing each tree
- eta: control the learning rate, i.e. scale the contribution of each tree by a factor of $0 < eta < 1$ when it is added to the current approximation
- gamma: minimum loss reduction required to make a further partition on a leaf node
- max_depth: maximum depth of a tree
- min_child_weight: minimum sum of instance weight (hessian) needed in a child
- nrounds: max number of boosting iterations
- subsample: subsample ratio of the training instance

Figure B.2 illustrates the outcomes of a grid search conducted across seven selected parameters of XGBOOST, in a similar way as in previous figure. The analysis indicates that no single model consistently outperforms others across different PPI sectors, underscoring the necessity for parameter tuning to optimize model fit. Additionally, parameter preferences exhibit no stability over time, with notable shifts observed when incorporating data from the post-COVID era.

Figure B.2: XGBoost Parameter Grid and Selection of the Optimal Model

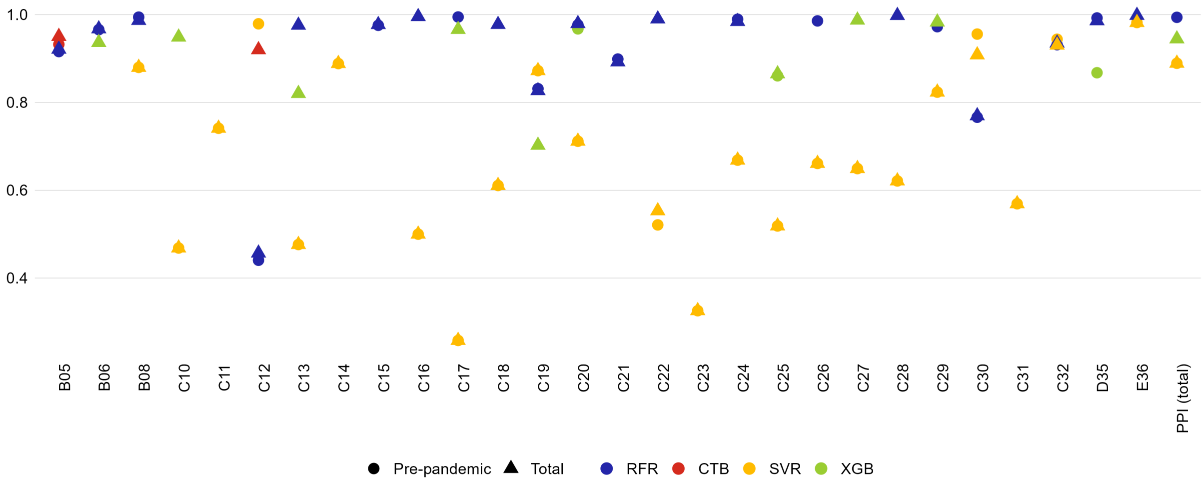


Note: Blue points show the results for pre-pandemic and red stand for values for the entire sample

Several key observations emerge from this exercise. When considering the entire period, which includes the more volatile and unstable post-COVID period, it is advisable to adjust specific parameters to better manage the increased data variability. Notably, increasing the subsample ratio of columns used in constructing each tree (colsample_by_tree) and the maximum depth of a tree (max_depth) are beneficial adjustments. Additionally, increasing the maximum number of boosting iterations (n_estimators) can significantly enhance the model's ability to handle more volatile datasets.

To assess the relevance of fine-tuning parameters across sectors, we compare the fine-tuned root mean squared errors for fixed time periods (pre-pandemic and total), to baseline RMSE. It is a simplification as conducting this exercise smoothly across different time periods would provide exact estimate of the necessity to fine-tune models. Eventually, some machine learning methods, such as penalized regressions, do perform continuous fine-tuning of the shrinkage parameter, either through BIC or time-series cross-validation. However, this approach is extremely demanding in terms of computational resources if more complex models are deployed, especially given our model setup with large number of sectors. Moreover, we anticipate that different methods will vary in their need for fine-tuning, a hypothesis that should still be evident from the comparison of the two periods we considered. A certain degree of simplification is necessary but still provides some important insights.

Figure B.3: Improvement in RMSE Compared to Baseline for Sector-based Tunes

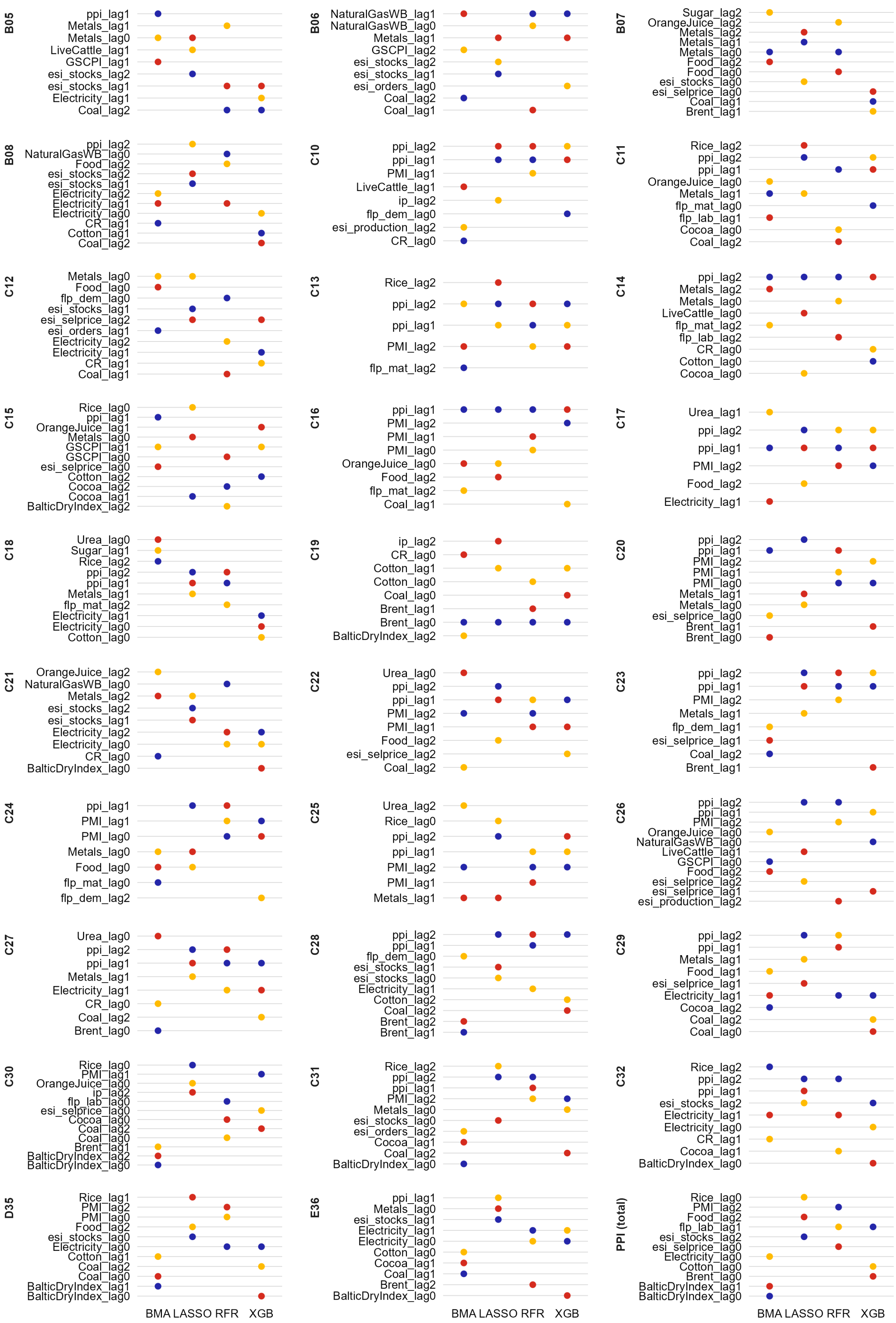


Note: The value stands for the share of fine-tuned model RMSE compared to the baseline setupe model estimated across a given time period, see more details in text.

Figure B.3 depicts the relativeRMSE of fine-tuned models compared to the baseline setup, highlighting instances where fine-tuning potentially improves performance (RMSE below 1). Among tree-based methods, such as XGB, CTB and RFR, the models generally perform well even without sector-specific fine-tuning. However, RFR shows a particular need for fine-tuning in the sector C12 (Manufacture of Tobacco Products), where it performs poorly in baseline estimates. XGB, while benefiting from fine-tuning in more sectors, shows more limited gains. Conversely, improvements due to fine-tuning in the case of CTB are noticeable only for two sectors. Support Vector Regression would benefit significantly from fine-tuning across virtually all sectors. In our baseline, SVR did not perform well, potentially due to the lack of continuous parameter fine-tuning akin to that employed in penalized regressions. In model description section 3.3, we noted strong similarities to penalized regressions, yet the direct impact of fine-tuning on forecasting performance has not been thoroughly discussed in the literature. We hypothesize that the generally poor performance of SVR, as reported in numerous studies, may relate to this phenomenon. Unfortunately, studies refrain from discussion on parameter tuning, leaving this issue open.

Appendix C: Additional Charts and Tables

Figure C.1: Ranking of the Three Best Variables (post-pandemic period)



Note: Sector code is on left, estimates for post-pandemic period, order given by colour - first (blue), second (red) and third (yellow).

Table C.1: Relative RMSEs for Three-month-ahead Forecasts—Baseline

			AR-based			Penalized regression					Tree-based				Neural nets		Factor/BMA			Hybrid OLS		Hybrid ARMAX		Hybrid NETS	
Sector	RW	AR	ARMA	ARMAX	RR	LASSO	ELNET	ALASSO	SVR	RFR	QRF	XGB	CTB	NNET	NNETAR	NNETARX	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX	RF-NNETARX	XGB-NNETARX
PPI (total)	1.08	0.96	1	0.84*	0.73***	0.74***	0.74***	0.74***	0.85	0.86**	0.9	0.82**	0.92	0.93	0.95	0.88**	0.71***	0.72***	0.74***	0.75**	0.74***	0.91	0.89	0.84**	0.82**
PPI (sector)	1.07	0.99	1	0.71***	0.71***	0.72***	0.7***	0.72***	0.85*	0.82***	0.86**	0.77***	0.94	0.92	1.75	0.83***	0.66***	0.67***	0.67***	0.64***	0.71***	0.81**	0.62***	0.85*	0.87*
B05	1.46	0.96*	1	1.21	0.95	0.95*	0.95*	0.95*	1.13	0.99	0.98	1.08	0.97	1.27	0.96*	1.05	1.12	1	0.99	1.12	1.16	1.3	1.17	0.99	1
B06	1.37	0.97	1	0.88	0.79***	0.73***	0.76***	0.73***	0.8**	0.79***	0.8***	0.82***	0.84***	0.93	1.45	0.89***	0.78***	0.71***	0.72***	0.76***	0.74***	0.87*	0.8**	0.87**	0.89**
B07	1.4	1.01	1	1.34	0.94	0.88*	0.89*	0.88*	1.39	0.93	0.92	0.89*	0.9*	1.38	0.96	0.96	1.05	0.88	0.89	1.05	0.94	0.98	1.07	0.97	0.97
B08	1.38	1.11	1	1.03	1.15	1.16	1.05	1.16	1.32	1.02	1.04	1.01	1.17	1.37	1.38	0.98	1.14	1.18	1.13	1.08	1.2	1.49	1.39	1.11	1.07
C10	1.03	1	1	0.93	0.76**	0.7***	0.66***	0.7***	1.21	0.8**	0.82**	0.72***	0.92	1.18	2.21	0.98	0.78**	0.67**	0.66***	0.68**	0.73**	0.68**	0.65**	0.99	0.92
C11	1.08	1.03	1	1.15	1.03	0.99	0.99	0.99	1.53	0.96	0.99	0.98	1.19	1.28	1.25	0.99	1.09	0.92*	0.96	1.04	1.16	0.99	1.11	1.07	1.07
C12	1.43	1.15	1	2.75	1.01	1	1	1	1.55	1.6	1.39	1.89	1.13	4.18	1.05	2.08	2.07	1.03	1.14	1.49	2.39	1.29	1.56	2.14	2.13
C13	1.02	1.06	1	1.14	1.07	1.02	1.02	1.02	1.61	0.81***	0.83***	0.92	1.16	1.36	0.93	0.97	0.96	0.91**	0.95	0.93	1.02	0.84***	0.97	1.14	1.13
C14	0.94	1.15	1	1.48	1.13	1.12	1.03	1.12	1.42	1.03	1.06	1.12	1.16	1.84	1.11	1.11	1.63	1.16	1.14	1.29	1.33	1.3	1.36	1.09	1.12
C15	1.19	1.04	1	1.68	0.97	0.99	0.96*	0.99	1.37	0.95*	0.95*	0.95	1.09	1.48	0.98	0.99	1.32	0.98	1.05	1.32	1.41	1.49	1.37	1	1
C16	1.16	1.15	1	0.98	0.74***	0.73***	0.76***	0.73***	0.96	0.82**	0.83**	0.83**	0.93	1.01	1.55	0.96	0.68***	0.73***	0.7***	0.73***	0.78***	0.55***	0.65***	0.99	1.05
C17	0.97	0.99**	1	1.04	0.75**	0.7***	0.76**	0.7***	1.35	0.97	0.99	0.86*	1.11	0.9	1.41	1.08	0.76**	0.61***	0.69***	0.65***	0.73**	0.74**	0.68***	1.08	1.12
C18	0.92	1.07	1	1.18	1.17	1.15	1.1	1.15	1.32	0.99	1.04	0.91*	1.14	1.22	0.92	1.05	1.1	1.12	1.14	1.11	1.16	1.2	1.32	1.12	1.1
C19	1.45	0.99*	1	0.46***	0.46***	0.43***	0.41***	0.43***	0.49***	0.54***	0.55***	0.62***	0.62***	0.77**	0.96**	0.63***	0.69**	0.42***	0.42***	0.4***	0.44***	0.43***	0.43***	0.57***	0.64***
C20	1.01	0.99	1	0.68***	0.55***	0.52***	0.51***	0.52***	0.71***	0.64***	0.66***	0.59***	0.71***	0.92	1.84	0.76***	0.52***	0.47***	0.48***	0.46***	0.5***	0.45***	0.5***	0.77***	0.79***
C21	1.12	0.95**	1	1.51	1.05	1.05	1.05	1.05	1.39	1.08	1.05	1.04	0.97	1.77	1	1.24	1.39	1.05	1.13	1.27	1.28	1.09	1.15	1.17	1
C22	0.91	0.96***	1	1.11	0.87*	0.77***	0.73***	0.77***	1.4	0.91*	0.93	0.92	1.01	1.25	1.86	1.08	0.84**	0.72***	0.74***	0.75***	0.82**	0.84**	0.88	1.07	1.09
C23	1.06	1.06	1	1.14	1	0.94*	0.95	0.94*	1.59	1.03	1.01	0.98	1.23	1.22	2.69	1.1	0.86**	0.87**	0.86**	0.87**	0.96	0.9	0.85*	1.18	1.13
C24	1.08	0.98***	1	0.75*	0.57***	0.52***	0.54***	0.52***	0.89	0.67***	0.71***	0.56***	0.69***	0.97	0.92***	0.69**	0.64**	0.49***	0.5***	0.54***	0.59***	0.58***	0.52***	0.68***	0.71***
C25	0.95	0.97	1	0.87	0.86	0.84	0.83	0.84	1.3	0.9	0.95	0.87	1.11	1.25	1.36	0.89	0.81	0.82	0.76*	0.81	0.91	0.83	0.8	0.92	0.94
C26	0.9	0.83***	1	1.77	1.13	1.11	1.15	1.11	1.3	1.18	1.16	1.07	0.87*	1.9	1.01	1.24	0.94	0.96	1	0.93	0.88	1.15	1.3	1.32	1.35
C27	1.02	1.06	1	1.2	0.99	0.99	0.94	0.99	1.52	0.92*	0.97	0.9	1.17	1.25	0.99	0.95	0.93	0.92	0.96	0.92	1.07	0.92	0.94	1.05	1.04
C28	1	0.98	1	1.25	1.13	1.1	1.1	1.1	1.74	1.08	1.16	1.11	1.45	1.55	2.34	0.97	1.07	1.01	1.05	0.95	1.23	0.99	1.07	1.16	1.13
C29	1.01	1.2	1	1.39	1.32	1.28	1.22	1.28	1.79	1.09	1.16	1.05	1.4	1.55	1.08	1.11	1.46	1.27	1.34	1.4	1.68	1.64	1.54	1.21	1.22
C30	1.29	1.18	1	1.58	1.04	1.05	1.06	1.05	1.78	1.16	1.11	1.39	1.19	2.09	1.06	1.15	1.22	1.08	1.02	1.31	1.49	1.31	1.43	1.14	1.25
C31	1.06	1.02	1	1.38	1.16	1.1	1.08	1.1	1.83	1.11	1.17	1.08	1.47	1.42	1.08	1.06	1.09	1.03	1.07	1.09	1.24	1.09	1.25	1.13	1.14
C32	1.33	1.06	1	1.17	1.01	1.01	0.99	1.01	1.19	1.04	1.03	1.02	1.13	1.54	1.12	1.01	1.13	1.04	1.03	1.14	1.27	1.21	1.2	1.02	1.06
D35	1.04	0.9	1	0.75**	0.78**	0.75**	0.76**	0.75**	0.79*	0.83**	0.84*	0.8**	0.89	0.87	2.15	0.83**	0.74**	0.74**	0.73**	0.7**	0.75**	0.98	0.76**	0.84*	0.85
E36	1.25	1.06	1	1.11	0.99	1	1	1	1.17	0.99	1	1.01	1.13	1.27	1.03	1.01	1.08	1	0.99	1.07	1.2	1.23	1.06	1.02	1.03

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: RW - Random Walk, AR - Autoregressive Model, ARMA - Autoregressive Moving Average, ARMAX - ARMA with Exogeneous Variables, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, ALASSO - Adaptive LASSO, SVR - Support Vector Regression, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, CATBOOST - Gradient Boosting on Decision Trees, NNET - Neural Network, NNETAR - Neural Network Autoregressive Model, NNETARX - NNETAR with Exogeneous Variables, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX, RF-NNETARX - Random Forest with Neural Nets, XGB-NNETARX - Gradient Boosting with Neural Nets

Table C.2: Relative RMSEs for Six-month-ahead Forecasts—Baseline

			AR-based		Penalized regression					Tree-based					Neural nets		Factor/BMA			Hybrid OLS		Hybrid ARMAX		Hybrid NETS	
Sector	RW	AR	ARMA	ARMAX	RR	LASSO	ELNET	ALASSO	SVR	RFR	QRF	XGB	CTB	NNET	NNETAR	NNETARX	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX	RF-NNETARX	XGB-NNETARX
PPI (total)	1.03	0.92**	1	0.76***	0.68***	0.69***	0.7***	0.69***	0.75***	0.8***	0.81***	0.78***	0.83***	0.88*	1	0.83***	0.67***	0.69***	0.67***	0.67***	0.71***	0.79**	0.74***	0.8***	0.78***
PPI (sector)	1.06	1.01	1	0.7***	0.67***	0.69***	0.68***	0.69***	0.76***	0.8***	0.82***	0.73***	0.86**	0.91	1.62	0.82***	0.63***	0.66***	0.63***	0.63***	0.66***	0.64***	0.57***	0.84***	0.84***
B05	1.3	1	1	1.36	1.01	1	1	1	1.21	0.99	0.98	1.03	1	1.33	0.99	1.02	1.19	1.05	1.04	1.27	1.23	1.26	1.66	1	1.02
B06	1.34	0.97*	1	0.92	0.79***	0.74***	0.76***	0.74***	0.84**	0.78***	0.79***	0.81***	0.84***	0.89*	1.58	0.86***	0.8***	0.71***	0.72***	0.8***	0.74***	0.89*	0.82***	0.88***	0.85***
B07	1.31	1.01	1	1.41	0.97	0.89*	0.91*	0.89*	1.48	0.92	0.92	0.94	0.93	1.38	1.01	0.93	1.05	0.89	0.9	1.05	1.09	0.93	1.24	1	0.98
B08	1.27	1.04	1	0.91*	1	1	0.95	1	1.16	0.92	0.95	0.94	1.04	1.09	1.38	0.96	0.98	1.02	0.98	0.98	1.09	1.29	1.28	0.97	0.98
C10	1.13	1.02	1	0.91	0.68**	0.61***	0.6***	0.61***	1.03	0.74***	0.75***	0.72***	0.84***	0.93	1.98	0.98	0.69**	0.58**	0.58**	0.59**	0.64**	0.6**	0.62**	0.96	0.94
C11	1.17	1.04	1	1.06	0.94	0.91*	0.91*	0.91*	1.37	0.88**	0.9**	0.9**	1.09	1.11	1.36	0.95	0.94	0.85**	0.86**	0.9*	0.99	0.81**	1.01	0.97	1
C12	1.6	1.16	1	3.39	1.04	1	1	1	1.6	1.65	1.36	2.67	1.24	4.11	1.04	2.26	2.33	1	1.19	1.72	2.45	1.53	1.44	1.96	1.61
C13	1	1.05	1	1.08	0.85***	0.82***	0.84***	0.82***	1.28	0.66***	0.68***	0.71***	0.93*	0.96	0.91	0.89***	0.8**	0.74***	0.76***	0.74***	0.83**	0.59***	0.81*	1.02	1.05
C14	0.99	1.11	1	1.26	1.05	1.03	0.96	1.03	1.24	0.97	0.99	0.99	1.09	1.68	1.04	1.03	1.56	1.02	1.05	1.09	1.24	1.25	1.31	1.09	1.02
C15	1.02	1.05	1	1.72	0.96	0.98	0.96*	0.98	1.26	0.94*	0.95*	0.91**	1.06	1.52	0.96*	0.97	1.34	0.97	1.03	1.24	1.45	1.46	1.57	0.99	1
C16	1.21	1.09	1	0.89**	0.67***	0.64***	0.65***	0.64***	0.84***	0.77***	0.78***	0.81***	0.84***	0.89*	1.57	0.83***	0.63***	0.62***	0.62***	0.62***	0.64***	0.71***	0.59***	0.85***	0.84***
C17	1	0.97***	1	0.97	0.55***	0.51***	0.56***	0.51***	0.92	0.74***	0.76***	0.68***	0.81**	0.78***	1.58	0.9*	0.53***	0.46***	0.49***	0.48***	0.49***	0.51***	0.47***	0.92	0.9*
C18	0.99	0.99	1	1.01	1.02	1.03	0.98	1.03	1.04	0.93**	0.97	0.91**	0.99	1.11	1.05	1	0.98	0.99	1.01	1	1.04	1.13	1.18	0.97	1
C19	1.39	0.99*	1	0.46***	0.45***	0.43***	0.41***	0.43***	0.46***	0.56***	0.56***	0.58***	0.64***	0.88	0.99	0.62***	0.72**	0.42***	0.42***	0.4***	0.43***	0.42***	0.41***	0.6***	0.65***
C20	1.02	0.98	1	0.64***	0.53***	0.49***	0.51***	0.49***	0.67***	0.62***	0.64***	0.56***	0.69***	0.92*	1.41	0.76***	0.52***	0.44***	0.46***	0.42***	0.5***	0.44***	0.43***	0.76***	0.75***
C21	1.09	0.94***	1	1.5	1.03	1.03	1.03	1.03	1.44	1.12	1.06	1.03	1.01	2.02	0.94	1.13	1.53	1.03	1.13	1.49	1.42	1.37	1.03	1.11	1.19
C22	1.04	1	1	1.16	0.77***	0.65***	0.61***	0.65***	1.22	0.78***	0.8***	0.72***	0.89***	1.01	1.76	0.99	0.73***	0.6***	0.64***	0.65***	0.72***	0.7***	0.65***	1.03	1.04
C23	1.06	1.08	1	1.07	0.9**	0.81***	0.84***	0.81***	1.27	0.96	0.93**	0.9***	1.09	1.15	2.02	1.11	0.74***	0.75***	0.73***	0.76***	0.81**	0.71**	0.72***	1.16	1.11
C24	1.12	0.98***	1	0.66***	0.53***	0.47***	0.5***	0.47***	0.81*	0.62***	0.65***	0.55***	0.65***	0.96	0.95	0.62***	0.62***	0.45***	0.46***	0.49***	0.57***	0.54***	0.5***	0.66***	0.65***
C25	0.98	0.98	1	0.84**	0.73**	0.7***	0.7***	0.7***	1.09	0.8**	0.84*	0.78**	0.99	1.04	1.38	0.87	0.68**	0.66***	0.64***	0.63***	0.73**	0.63***	0.72**	0.85	0.83*
C26	0.84	0.77***	1	1.6	0.97	0.93**	0.97	0.93**	1.05	1.06	1.02	0.86**	0.75***	1.57	0.96	1.22	0.82**	0.8***	0.84***	0.84***	0.82*	1.15	1.25	1.36	1.34
C27	1.04	1.08	1	1.12	0.88*	0.9*	0.84**	0.9*	1.29	0.85***	0.9*	0.77***	1.06	1.1	0.91**	0.94	0.84**	0.83**	0.87**	0.88*	0.9	0.83**	0.8***	0.99	1
C28	0.86	0.94*	1	1.09	0.92	0.93	0.94	0.93	1.37	0.91	0.96	0.89	1.2	1.32	1.65	0.92	0.89	0.83**	0.86*	0.79**	0.99	0.73**	0.86	0.98	0.97
C29	1.02	1.16	1	1.2	1.1	1.1	1.04	1.1	1.51	0.96	1.01	0.91	1.21	1.27	1.09	0.99	1.23	1.1	1.13	1.23	1.39	1.48	1.26	1.06	0.99
C30	1.16	1.2	1	1.55	0.99	1.01	1.02	1.01	1.75	1.08	1.06	1.34	1.13	2.73	1.01	1.08	1.18	1.04	0.98	1.29	1.46	1.26	1.45	1.19	1.18
C31	0.94	0.98	1	1.23	1.04	1.01	1	1.01	1.6	1.05	1.09	1.07	1.31	1.25	1.1	1.11	0.98	0.92	0.97	0.94	1.11	0.95	1.13	1.11	1.09
C32	1.22	1.08	1	1.15	1	1.01	0.99	1.01	1.28	1.05	1.02	0.95	1.12	1.52	1.01	1.05	1.09	1.01	1.02	1.14	1.24	1.14	1.18	1.08	1.02
D35	1.14	0.94	1	0.78**	0.79**	0.78**	0.8**	0.78**	0.79**	0.85*	0.86*	0.8**	0.87*	0.88	2.14	0.84**	0.74**	0.78**	0.74**	0.75**	0.75**	0.84*	0.75**	0.82**	0.83**
E36	1.34	1.07	1	1.11	1	1	1	1	1.17	0.99	0.99	1.01	1.15	1.4	1.03	0.99	1.06	1	1	1.11	1.22	1.3	1.3	1.03	1.01

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: RW - Random Walk, AR - Autoregressive Model, ARMA - Autoregressive Moving Average, ARMAX - ARMA with Exogeneous Variables, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, ALASSO - Adaptive LASSO, SVR - Support Vector Regression, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, CATBOOST - Gradient Boosting on Decision Trees, NNET - Neural Network, NNETAR - Neural Network Autoregressive Model, NNETARX - NNETAR with Exogeneous Variables, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX, RF-NNETARX - Random Forest with Neural Nets, XGB-NNETARX - Gradient Boosting with Neural Nets

Table C.3: Relative RMSEs for Three-month-ahead Forecasts—Pipeline Pressures

	AR-based		Penalized regression			Tree-based			Factor/BMA		Hybrid OLS			Hybrid ARMAX	
Sector	AR	ARMA	RR	LASSO	ELNET	RFR	QRF	XGB	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX
PPI (total)	0.96	1	0.74***	0.74***	0.74***	0.87**	0.94	0.82**	0.71***	27.74	0.74***	0.74***	0.79*	0.85*	0.93
PPI (sector)	0.99	1	0.71***	0.73***	0.73***	0.83***	0.89*	0.83**	0.7***	21.64	0.67***	0.68***	0.84**	0.63***	0.73***
B05	0.96*	1	0.97	0.96	0.95	1.02	0.97	1.17	1.2	7.28	1.04	1.13	1.09	1.44	1.16
B06	0.97	1	0.81***	0.73***	0.74***	0.82***	0.83***	0.83***	0.85**	1.17	0.73***	0.78**	0.87**	0.77***	0.76***
B07	1.01	1	1.02	0.89*	0.89*	0.95	0.91*	0.99	1.65	2.96	0.92	1.05	1.11	0.97	0.97
B08	1.11	1	1.05	1.07	1.08	1	1.08	0.98	1.01	328.37	1.03	1.08	1.12	1.1	1.05
C10	1	1	0.66***	0.7***	0.72***	0.83**	0.84**	0.82**	0.75*	119.34	0.64***	0.74**	0.78*	0.81*	0.81
C11	1.03	1	0.93	0.88**	0.88**	0.87**	0.91*	0.98	1.13	40.29	0.9*	1	1.13	0.91	0.97
C12	1.15	1	1.06	1	1	1.71	1.05	3.06	3.37	28.64	1.2	1.93	1.72	1.99	1.79
C13	1.06	1	0.87***	0.95	0.96	0.82***	0.87**	0.91	0.93	39.25	0.82***	0.93	0.87*	0.89	0.93
C14	1.15	1	1.04	1.08	1.09	0.92*	0.98	1.12	1.53	47.29	0.97	1.29	1.12	1.09	1.18
C15	1.04	1	0.91***	0.96*	0.97	0.87***	0.88***	0.89*	1.45	53.37	0.91*	1.24	1.07	1.27	1.3
C16	1.15	1	0.8**	0.73***	0.74***	0.85*	0.88	1.04	0.85	9.38	0.72***	0.73***	0.76**	0.98	1.18
C17	0.99**	1	0.72***	0.71***	0.72***	1.01	1.01	0.85*	0.75**	41.19	0.68***	0.68***	0.83*	0.78**	0.77**
C18	1.07	1	0.96	1.05	1.03	0.88**	0.99	0.84**	0.99	22.23	0.92*	1.03	0.95	1	1.06
C19	0.99*	1	0.52***	0.44***	0.44***	0.61***	0.61***	0.62***	0.72***	1.96	0.41***	0.38***	0.82*	0.38***	0.42***
C20	0.99	1	0.57***	0.51***	0.52***	0.66***	0.71***	0.68***	0.6***	7.57	0.49***	0.48***	0.58***	0.5***	0.52***
C21	0.95**	1	1.02	1.05	1.05	1.1	0.96	0.99	1.75	61.25	1.08	1.27	1.34	1.25	1.04
C22	0.96***	1	0.71***	0.74***	0.73***	0.78***	0.81***	0.75***	0.89	35.38	0.69***	0.73***	0.8**	0.76***	0.66***
C23	1.06	1	0.84***	0.93**	0.91***	1	0.97	0.98	0.8**	21.5	0.8***	0.83***	0.95	0.89**	0.89*
C24	0.98***	1	0.61***	0.53***	0.53***	0.73***	0.77**	0.67***	0.79	7.33	0.51***	0.5***	0.87	0.54***	0.6***
C25	0.97	1	0.79*	0.92	0.93	0.91	0.96	0.83	0.86	517.23	0.8*	0.84	0.99	0.92	0.89
C26	0.83***	1	1.33	1.12	1.16	1.17	1.04	0.95	1.45	43.15	1.19	1.38	1.37	1.26	1.18
C27	1.06	1	0.84**	0.95	0.96	0.84***	0.91*	0.8***	1	75.01	0.83**	0.88	1.16	0.89	0.91
C28	0.98	1	0.89*	0.9**	0.91**	0.88***	0.96	0.88**	1.07	104.47	0.86**	1.04	1.21	0.97	1.02
C29	1.2	1	1.11	1.1	1.1	0.92**	1.09	0.91*	1.34	158.46	1.04	1.34	1.09	1.22	1.29
C30	1.18	1	0.96	0.9**	0.94*	0.92	0.92**	1.25	1.91	60.5	1.16	1.15	1.4	1.67	1.17
C31	1.02	1	0.98	1.01	1.02	0.96	1.06	0.89*	1.12	337.25	0.94	0.99	1.17	0.98	1.01
C32	1.06	1	0.99	1.05	1.01	0.96	0.93**	1	1.34	236.33	0.97	1.27	1.23	1.05	1.19
D35	0.9	1	0.8**	0.77**	0.77**	0.87	0.9	0.93	0.79*	2.23	0.74**	0.74**	0.92	0.7**	0.83*
E36	1.06	1	0.99*	1	1	0.97	0.98	1.04	1.09	903.86	0.98	1.05	1.17	1.04	1

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: AR - Autoregressive Model, ARMA - Autoregressive Moving Average, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX

Table C.4: Relative RMSEs for Six-month-ahead Forecasts—Pipeline Pressures

	AR-based		Penalized regression			Tree-based			Factor/BMA		Hybrid OLS			Hybrid ARMAX	
Sector	AR	ARMA	RR	LASSO	ELNET	RFR	QRF	XGB	FC	BMA	RF-OLS	XGB-OLS	CTB-OLS	RF-ARMAX	XGB-ARMAX
PPI (total)	0.92**	1	0.68***	0.69***	0.69***	0.8***	0.83***	0.77***	0.67***	18.53	0.67***	0.67***	0.75**	0.73***	0.74***
PPI (sector)	1.01	1	0.67***	0.69***	0.69***	0.83***	0.82***	0.85*	0.65***	18.6	0.62***	0.59***	0.74***	0.58***	0.64***
B05	1	1	1.05	0.99	0.99	0.98	1	1.13	1.18	7.94	1.08	1.21	1.06	1.32	1.31
B06	0.97*	1	0.8***	0.76***	0.75***	0.82***	0.82***	0.79***	0.87**	1.13	0.72***	0.69***	0.91	0.75***	0.74***
B07	1.01	1	1.05	0.9*	0.9*	0.97	0.93	0.98	1.62	2.8	0.93	1.05	1.17	0.95	0.99
B08	1.04	1	0.93	0.95	0.94	0.9*	0.95	0.91*	0.9*	219.94	0.9	0.89*	0.97	0.95	0.96
C10	1.02	1	0.58**	0.61***	0.62***	0.76***	0.79***	0.73***	0.68*	88.22	0.56***	0.62**	0.66**	0.66**	0.64**
C11	1.04	1	0.85***	0.8***	0.81***	0.79***	0.83***	0.83**	1.01	32.89	0.8***	0.91	0.89	0.83*	0.84*
C12	1.16	1	1.09	1	1	1.86	1.05	3.09	4.18	27.86	1.23	2.35	2.14	2.08	1.88
C13	1.05	1	0.69***	0.77***	0.8***	0.69***	0.71***	0.72***	0.7***	26.7	0.64***	0.69***	0.7***	0.72***	0.9
C14	1.11	1	0.99	1.03	1.03	0.91**	0.95	1.09	1.56	42.42	0.93	1.14	1.19	1.05	0.97
C15	1.05	1	0.92***	0.96*	0.96*	0.88***	0.89***	0.9**	1.44	32.05	0.9*	1.05	0.98	1.16	1.2
C16	1.09	1	0.7***	0.64***	0.65***	0.78***	0.79***	0.98	0.73**	8.2	0.63***	0.63***	0.64***	0.94	0.9
C17	0.97***	1	0.55***	0.52***	0.53***	0.79***	0.82***	0.66***	0.57***	27.5	0.5***	0.51***	0.62***	0.59***	0.61***
C18	0.99	1	0.86***	0.95	0.95	0.86***	0.91**	0.77***	0.89*	16.69	0.82***	0.91	0.93	0.89	0.86*
C19	0.99*	1	0.57***	0.45***	0.46***	0.63***	0.61***	0.65***	0.79**	1.93	0.42***	0.41***	0.83	0.39***	0.41***
C20	0.98	1	0.55***	0.47***	0.5***	0.65***	0.7***	0.62***	0.59***	7.73	0.47***	0.39***	0.6***	0.48***	0.48***
C21	0.94***	1	1.03	1.03	1.03	1.22	0.97	0.95	1.8	61.28	1.14	1.35	1.39	1.27	1.34
C22	1	1	0.62***	0.66***	0.64***	0.68***	0.71***	0.67***	0.81*	38.97	0.58***	0.57***	0.66***	0.6***	0.65***
C23	1.08	1	0.75***	0.81***	0.82***	0.94**	0.9***	0.91***	0.68***	19.02	0.7***	0.73***	0.79**	0.77***	0.82***
C24	0.98***	1	0.53***	0.48***	0.48***	0.66***	0.7***	0.61***	0.71**	6.68	0.47***	0.5***	0.68***	0.49***	0.5***
C25	0.98	1	0.67***	0.74**	0.76**	0.81**	0.83*	0.68***	0.81**	492.09	0.68***	0.74**	0.77**	0.83*	0.73**
C26	0.77***	1	1.28	0.96	0.99	1.14	0.99	0.88**	1.35	40.45	1.12	1.33	1.16	1.16	0.99
C27	1.08	1	0.74***	0.89**	0.91*	0.79***	0.85**	0.64***	0.93	31.2	0.73***	0.83**	0.9	0.79**	0.76**
C28	0.94*	1	0.7***	0.75**	0.76**	0.77***	0.86*	0.72***	0.85	67.41	0.67***	0.77**	0.94	0.72**	0.76**
C29	1.16	1	0.96	0.98	0.98	0.85***	0.96	0.88*	1.21	89.04	0.88**	1.01	1.04	1.06	1.05
C30	1.2	1	0.95	0.86**	0.96	0.87*	0.85***	1.16	2.07	69.04	1.21	1.39	1.33	1.69	1.51
C31	0.98	1	0.89*	0.92	0.93	0.89**	0.97	0.83***	1.02	216.1	0.86**	0.91	1.07	0.96	0.88*
C32	1.08	1	0.97	0.98	1	0.96*	0.95*	0.91	1.2	248.63	0.95	1.15	1.28	1.01	1.23
D35	0.94	1	0.78**	0.77**	0.76**	0.89	0.85*	0.98	0.75**	1.89	0.73**	0.68***	0.9	0.73**	0.78**
E36	1.07	1	1	1	1	0.99	1	1.01	1.12	698.53	0.99	1.08	1.21	1.07	0.98

Note: The three lowest values for each sector are highlighted in red, expressed as a share of the ARMA RMSE. ***, **, and * indicate rejection at the 1%, 5%, and 10% levels in the one-tailed Diebold-Mariano test. Methods used: AR - Autoregressive Model, ARMA - Autoregressive Moving Average, RIDGE - Ridge Regression, LASSO - Lasso, ELNET - Elastic Net, RF - Random Forest, QRF - Quantile Regression Forest, XGB - Extreme Gradient Boosting, FC - Factor Model, BMA - Bayesian Model Averaging, RF-OLS - Random Forest with OLS, XGB-OLS - Extreme Gradient Boosting with OLS, CatBoost - Gradient Boosting on Decision Trees with OLS, RF-ARMAX - Random Forest with ARMAX, XGB-ARMAX - Extreme Gradient Boosting with ARMAX

CNB Working Paper Series (since 2024)

WP 2/2025	Soňa Benecká	*Forecasting disaggregated producer prices: A fusion of machine learning and econometric techniques*
WP 1/2025	Martin Hodula Lukáš Pfeifer	*Payment holidays, credit risk, and borrower-based limits: Insights from the Czech mortgage market*
WP 9/2024	Jaromír Baxa Tomáš Šestořád	*Economic policy uncertainty in Europe: Spillovers and common shocks*
WP 8/2024	Martin Hodula Jan Janků Simona Malovaná Ngoc Anh Ngo	*Geopolitical risks and their impact on global macro-financial stability: Literature and measurements*
WP 7/2024	Jan Brůha Adam Ruschka Jan Šolc	*LUCI: Your best friend for measuring labor market tightness*
WP 6/2024	Aleš Bulíř Jan Vlček	*The mirage of falling R-stars*
WP 5/2024	Natálie Dvořáková Tomáš Šestořád	*Origins of post-COVID-19 inflation in central European countries*
WP 4/2024	Volha Audzei Ivan Sutóris	*A heterogeneous agent model of energy consumption and energy conservation*
WP 3/2024	František Brázdik Tatiana Keseliová Karel Musil Radek Šnobl Jan Šolc Stanislav Tvrz Jan Žáček	*Understanding inflation expectations: Data, drivers and policy implications*
WP 2/2024	Jiří Gregor	*A stress test approach to the calibration of borrower-based measures: A case study of the Czech Republic*
WP 1/2024	Michal Franta Jan Vlček	*Wage-price spirals: A risk-based approach*

CZECH NATIONAL BANK
Na Příkopě 28
115 03 Praha 1
Czech Republic

ECONOMIC RESEARCH DIVISION
Tel.: +420 224 412 321
Fax: +420 224 412 329
http://www.cnb.cz
e-mail: research@cnb.cz

ISSN 1803-7070