WIFO

Mit Unterstützung von Bund und dafne.at

Bundesministerium
Land- und Forstwirtschaft,
Regionen und Wasserwirtschaft

DaFNE

Robuste Versorgungs-ketten in der Agrar- und Nahrungsmittelwirtschaft

Forecasting Food Price Inflation in Austria

Josef Baumgartner, Serguei Kaniovski (WIFO)

Wissenschaftliche Assistenz: Astrid Czaloun, Ursula Glauninger, Christine Kaufmann

Juli 2023
Österreichisches Institut für Wirtschaftsforschung

Robuste Versorgungs-ketten in der Agrar- und Nahrungsmittelwirtschaft

Forecasting Food Price Inflation in Austria

Josef Baumgartner, Serguei Kaniovski (WIFO)

Juli 2023

Österreichisches Institut für Wirtschaftsforschung,
Im Auftrag des Bundesministeriums für Land- und Forstwirtschaft, Regionen und Wasserwirtschaft (BML)

Wissenschaftliche Assistenz: Astrid Czaloun, Ursula Glauninger, Christine Kaufmann

Uncertainty about the persistence of the recent rise in inflation poses significant challenges for practitioners and policymakers. Food price dynamics have emerged as an important contributor to headline inflation, raising concerns about distributional effects and implications for competition policy. The study aims to provide reliable forecasts for five broad categories of food prices in Austria. To this end, we evaluate a diverse set of empirical models, review their ability to predict inflation based on available leading indicators, and discuss the current outlook. From an academic perspective, we examine the extent to which the forecasting accuracy has deteriorated compared to the pre-pandemic period and discuss the relative performance of time series models, regression trees, and machine learning approaches in the pre-pandemic and post-pandemic periods.

Forecasting food price inflation in Austria

Josef Baumgartner * Serguei Kaniovski †

July 31, 2023

Abstract

Uncertainty about the persistence of the recent rise in inflation poses significant challenges for practitioners and policymakers. Food price dynamics have emerged as an important contributor to headline inflation, raising concerns about distributional effects and implications for competition policy. The study aims to provide reliable forecasts for five broad categories of food prices in Austria. To this end, we evaluate a diverse set of empirical models, review their ability to predict inflation based on available leading indicators, and discuss the current outlook. From an academic perspective, we examine the extent to which the forecasting accuracy has deteriorated compared to the pre-pandemic period and discuss the relative performance of time series models, regression trees, and machine learning approaches in the pre-pandemic and post-pandemic periods.

JEL-Codes: E37, L66

Key Words: food price inflation, model selection, forecasting, time series models, regression trees, machine learning

*Austrian Institute of Economic Research (WIFO),
Arsenal, Object 20, 1030 Vienna, Austria.
Phone: +431 / 798 26 01 - 230
Fax: +431 / 798 93 86
E-mail: josef.baumgartner@wifo.ac.at

†Austrian Institute of Economic Research (WIFO),
Arsenal, Object 20, 1030 Vienna, Austria.
Phone: +431 / 798 26 01 - 231
Fax: +431 / 798 93 86
Email: serguei.kaniovski@wifo.ac.at

Contents

List of Tables

List of Figures

1 Introduction

Food is a vital commodity for which there are no substitutes. The lack of substitutes for food makes its demand unresponsive to changes in price, with price increases having a significant impact on income. Rising food prices lead to a reallocation of household spending, reducing disposable income available for non-essential goods and services, such as luxury goods, vacations, or entertainment, affecting overall consumer demand and potentially also economic growth. On the other hand, the additional revenue generated by higher prices flows back to producers and traders in the food supply chain, increasing their revenues and encouraging investment. The adverse impact of food inflation on economic growth may therefore be less in the long run than in the short run.

Over the last 70 years, the share of food (incl. non-alcoholic beverages (COICOP CP01)) in the total expenditure of Austrian consumers has fallen from around 45 percent to currently (household budget survey 2019/2020) around 12 percent. This decline reflects on the one hand, productivity progress in agriculture and the food industry, and on the other hand, the rise in living standards and thus higher spending on non-food items and services. These patterns are consistent with those in other industrialized countries (Whitmore Schanzenbach, Nunn, Bauer and Mumford 2016). Consumers in developing countries spend a significantly higher proportion of their income on food Meade and Rosen (1996), but this proportion tends to decrease as living standards rise.

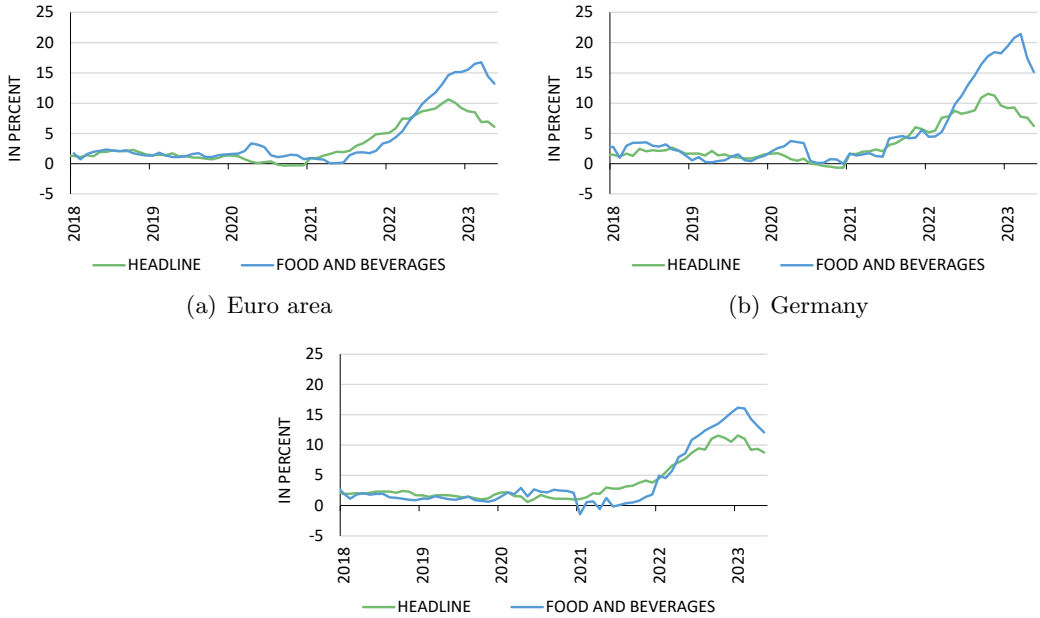
Large increases in food prices affect low-income households relatively more than higher-income households. The share of expenditure on food and beverages depends on the level of household income: In the first decile with the lowest income, it accounts for about $17\frac{1}{2}$ percent of total consumption expenditure, according to the 2019/2020 Household Budget Survey, while in the tenth decile with the highest income it is $13\frac{1}{4}$ percent.

In its quarterly economic forecast, WIFO forecasts an aggregate for food that includes non-alcoholic and alcoholic beverages and tobacco products as a single item (Baumgartner 2022). This project disaggregates this forecast to five items for the forecast horizon of twelve months. The food items are divided into three categories: (i) pure or predominantly plant-based foods, (ii) animal-based foods, and (iii) a miscellaneous category of foods that do not clearly fall into one of the previous two categories. In addition, the prices of beverages are divided into (iv) non-alcoholic beverages and (v) alcoholic beverages.

Compared to other consumer goods, the food value chain is significantly shorter and more dependent on regional developments (production conditions, weather, climate). Changes in agricultural producer prices in Austria and the EU (especially in Italy, Germany, Spain, and the Netherlands) have a more direct impact on domestic consumer prices for food. This is also reflected in the selection of variables leading to an improvement in the food inflation forecast.

Inflation rates have increased substantially after the severe economic downturn caused by the pandemic and the subsequent rise in energy prices (especially in Europe) following the outbreak of the war between Russia and Ukraine (Baumgartner and Sinabell 2021). Food prices rose at an above-average rate, i.e. more than the CPI (HICP) as a total. This development was also observed in the other Euro area countries, although food price inflation in Austria was somewhat weaker than in Germany and the average of the Euro area countries (see Figures 1 and 2). The main reason for this difference in the last two years is the higher price increases for plant-based and animal-based food products in Germany. Prices for non-alcoholic beverages have risen more sharply in Austria but did not have a strong impact because of the lower weight of this position as compared to food.

Figure 1: Food and beverage inflation compared with headline inflation



(a) Euro area

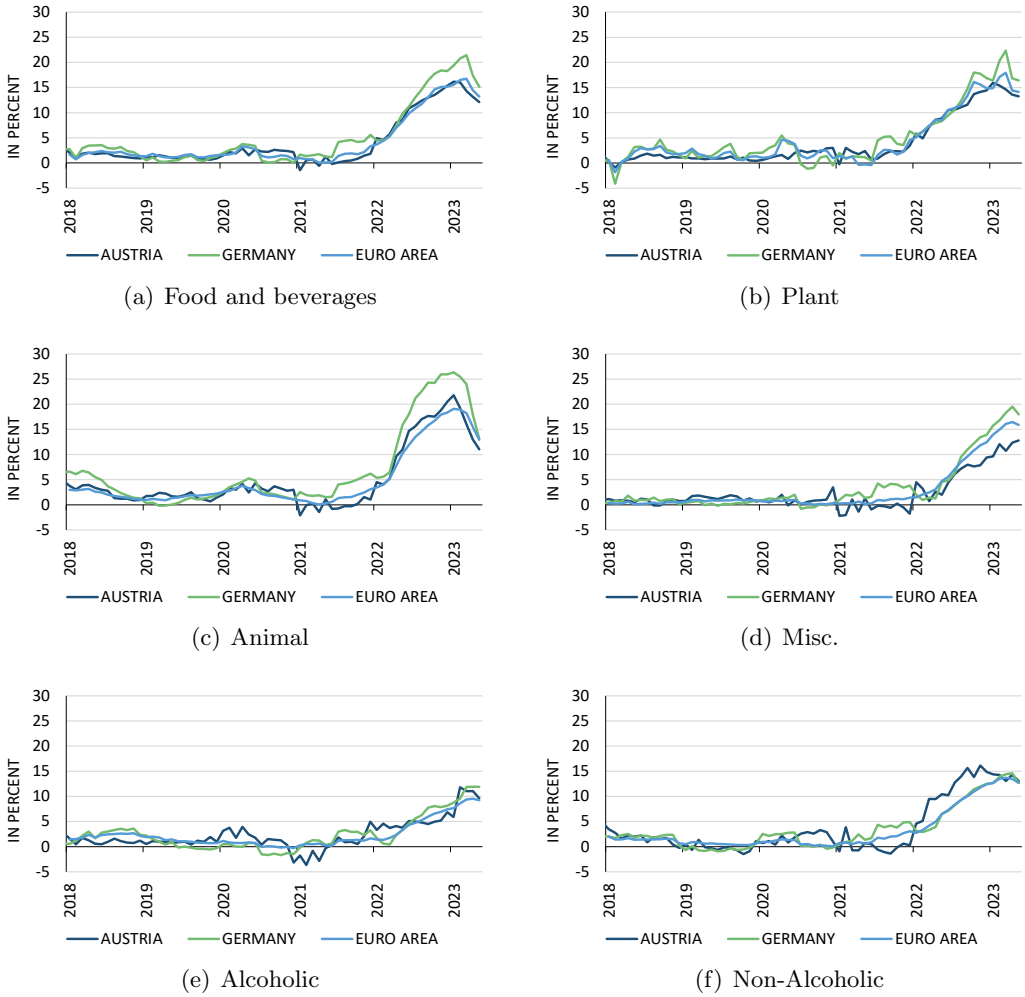
(b) Germany

(c) Austria

The figure compares headline inflation rates with food and beverage inflation rates in Austria, Germany, and the Euro area. The rates of headline inflation are broadly comparable across the three economies, although in Austria headline inflation has been more persistent. In Austria the food and beverage inflation has not increased as much relative to the headline inflation as in the Euro area particularly in Germany.

The study aims to provide reliable forecasts for five broad categories of food prices in Austria. We evaluate a diverse set of empirical models, review their ability to predict inflation based on available leading indicators, and discuss the current outlook, and examine the extent to which the forecasting accuracy has deteriorated compared to the pre-pandemic period and discuss the relative performance of time series models, regression trees, and machine learning approaches in the pre-pandemic and post-pandemic periods. Following this introduction, Section 2 summarizes the time series data used to estimate and evaluate the competing forecast models presented in Section 3. Section 4 then compares the overall as well as relative forecasting performance of the models and Section 5 discusses the current forecast for the period June 2023 to May 2024. The final section offers concluding remarks.

Figure 2: Food and beverage inflation rates in Austria, Germany and the Euro area



(a) Food and beverages

(b) Plant

(c) Animal

(d) Misc.

(e) Alcoholic

(f) Non-Alcoholic

The figure compares the inflation rates of the food and beverages aggregate and its five subcategories for Austria, Germany and the Euro area. Inflation rates for plant-based and animal-based food products increased more than those for beverages, especially in Germany. Austria experienced a relatively strong increase in inflation rates for non-alcoholic beverages.

2 Data

The data include a set of monthly economic indicators used to forecast each of the five target time series. The target series are composite price indices obtained by weighting appropriate CPI components:

1. PLANT-based food products (62 CPI components);
2. ANIMAL-based food products (45 CPI components);
3. MISCellaneous food products (11 CPI components);
4. ALCOHOLIC beverages (10 CPI components);
5. NON-ALCOHOLIC beverages (12 CPI components).

Appendix A lists the individual components of the CPI, their weights in the current consumption basket and their aggregation weights in the target series. The aggregation weights are expressed in percent and sum to unity. The sum of the consumption basket weights is less than one because the basket contains many other goods and services that are not included in the target series. Miscellaneous food products bundles those components that cannot be clearly classified as plant or animal or clearly do not belong to either category, e.g. salt or deep-frozen convenience food.

Table 1: Groups of indicators.

	AT	CH	EA	EU	US	World	Total
Production				2			2
Trade				2			2
Supply chains						1	1
Commodity prices				1	3	58	62
Import prices	10		7				17
Exchange rates					1		1
Producer prices	22	11	97	62			192
Wholesale prices	16		1				17
CPI	9						9
Wages	4						4
Financial			3		2		5
Sentiment			3	2			5
Total	61	11	111	69	6	59	317

Food price inflation is shaped by a complex interplay of numerous determinants (Baumgartner and Sinabell 2021). The empirical approach taken in this study is to collect a comprehensive data set on potential determinants of food price inflation in Austria and to allow a model to identify those indicators that improve the forecasting performance. Table 1 lists the main groups of indicators. The data set covers import prices, producer prices, wholesale prices, and consumer

prices of major trading partners, as well as a number of relevant commodity prices, including fuels (Baumgartner and Sinabell 2021).

Production and trade: The general level of economic activity and demand is an important determinant of overall inflationary pressures and, in particular, of food price inflation. Increased output levels can result in surplus supply, leading to lower prices. Conversely, reduced output due to economic downturns or supply chain disruptions can lead to temporary shortages and higher prices. International trade can also impact food prices, as imports and exports influence supply and demand dynamics in domestic markets. We capture the business cycle using monthly production indices for the primary sector, manufacturing, and energy in the EU 27.

Supply chains play a pivotal role in the availability and affordability of food. Supply chain disruptions due to political crises, or transportation bottlenecks can affect (timely) delivery of agricultural commodities and finished products. An index of the situation with supply chains provides insights into the efficiency and resilience of these systems, impacting food price inflation.

Commodity prices: Food price movements are closely linked to the prices of key commodities, such as grains, oilseeds, coffee, sugar, and livestock. Fluctuations in commodity prices can be influenced by weather conditions, geopolitical events, and global supply and demand imbalances. They propagate through food supply chains, affecting input costs and ultimately influencing consumer prices.

Import prices and exchange rate: Import prices are influenced by a combination of factors, including international commodity prices, exchange rates, transportation costs, tariffs, and trade policies. Higher import prices can translate into increased consumer prices, particularly in economies heavily reliant on imports. Fluctuations in exchange rates can increase the cost of imported food products, leading to higher prices for consumers, whereas a strengthening currency can have the opposite effect on food price inflation.

Producer and wholesale prices reflect the costs incurred by food producers. Higher input costs, such as labor, energy, and raw materials, are often passed on to consumers, contributing to overall food price inflation. This is the largest group of indicators used to forecast the target series, with about sixty percent of all indicators belonging to this category. The extent to which costs are passed on to consumers depends on prevailing market conditions and the intensity of competition in the relevant markets. Wholesale prices serve as an intermediary between producer and consumer prices that can be driven by factors such as production levels, market demand, transportation costs, and supply chain efficiencies. Fluctuations in wholesale prices can influence the pricing decisions made by retailers, impacting the final prices paid by consumers.

CPI: The target series comprise highly disaggregated CPI indicators weighted according to their relative importance. The set of indicators used to forecast each target series includes further CPI aggregates related to energy, different kinds of fuels and other agricultural products used as inputs to food production. Energy prices appear to be a major factor behind food price inflation.

Wages are a potentially important determinant of food price inflation in the medium term. Higher wages raise production costs, which can be passed on to consumers through higher prices. The role of wage dynamics is particularly relevant for Austria, with its highly centralized wage bargaining system (collective bargaining rate coverage of 98 percent) in which representatives of employers and trade unions negotiate minimum hourly wage increases and other working conditions. Conversely, stagnant or falling real wages have less of an impact on production costs, but also constrain purchasing power and demand, which can lead to lower food price increases.

Financial market conditions can have an indirect impact on food price inflation, as many

agricultural commodities such as coffee, corn, and sugar are traded internationally and have become an increasingly popular alternative to traditional financial investments. Stock market volatility can have a direct impact on commodity markets and consumer confidence, which in turn can affect consumption patterns and food price dynamics.

Sentiment indicators reflect the current economic conditions and near future expectations of consumers and producers. Positive sentiment and optimistic economic expectations can drive spending, increase demand and potentially lead to higher prices. Pessimistic sentiment can dampen consumer and producer confidence and constrain demand, potentially mitigating upward pressure on food prices.

2.1 Seasonality and stationarity

The majority of the time series used to estimate the forecasting models is expressed in year-over-year (yoy) growth rates. The choice of growth rates over the previous year's month rather than over the previous month is motivated by a desire to minimize the impact of seasonality on model estimates. This approach is validated by the ARIMA models discussed below, which indeed show the absence of seasonal terms in the yoy growth rates. The exception to this convention is sentiment indicators, which may turn negative if pessimism outweighs optimism, and are therefore included in levels.

The sample used in model selection and validation covers the period from January 2001 to December 2022 for a total of 264 monthly observations. The forecast horizon extends 12 months into the future, i.e., currently from June 2023 to May 2024. Models are tested based on hypothetical out-of-sample forecasts for a given year (2018, 2019, 2020, 2021), as explained in more detail below. In terms of geographic coverage, about sixty percent of all indicators reflect conditions in European markets, as these are most relevant to Austria.

The growth rates are weakly stationary. The KPSS test shows that the null hypothesis of level stationarity or trend stationarity cannot be rejected for the five target series at all standard statistical significance levels (Kwiatkowski, Phillips, Schmidt and Shin 1992). Only 7.2 percent and 18.1 percent of all indicators fail the test at the 5 percent significance level for level and trend stationarity, respectively, but pass both tests at the 1 percent level. The tests are based on 15 lagged values of the time series, as is appropriate for monthly data.

3 Forecasting models

Understanding the causes of inflation as a theoretical endeavor and forecasting inflation as a practical application have a long tradition in economics due to the monetary policy objectives pursued by central banks, and there is a large body of literature on the subject (Faust and Wright 2013). Early works on inflation forecasting, such as Stock and Watson (1999), laid the foundation for research in this area. Since then, numerous studies have investigated into various factors contributing to inflation, including expectations, fluctuations in financial markets, supply shocks emanating from natural disasters, climate change and international supply chains. The topic of forecasting food inflation is particularly prevalent in agricultural and resource economics, where price developments are linked to the natural conditions that determine agricultural yields. Government bodies such as US Department of Agriculture (Kuhns, Leibtag, Volpe and Roeger 2015) and international organizations such as the World Bank (for Reconstruction and (IBRD 2023) provide regular forecasts for internationally traded commodities and retail food prices.

Inflation forecasting employs a variety of models, ranging from univariate and multivariate time series models, as demonstrated by Ahumada and Cornejo (2016), to state-space models and mixed-frequency models that can handle mixed-frequency data, e.g. Modugno (2013) and Monteforte and Moretti (2013), and increasingly more often regression trees and machine learning models that are particularly well-suited in data rich environments featuring large quantities of high-frequency price data. As technology and data availability have advanced, new forecasting models incorporating more advanced techniques have emerged. Recent developments in inflation forecasting have seen the utilization of large volumes of such data made available through internet commerce, e.g. Modugno (2013), Cavallo and Rigobon (2016), Gorodnichenko and Talavera (2017), Cavallo (2018), Aparicio and Bertolotto (2020), Macias, Stelmasiak and Szafranek (2023), and consumer survey data and internet search data (Jo and Lusk 2016). Analyzing such extensive data sets requires the application of new forecasting models, often incorporating elements of machine learning (e.g., Menculini, Marini, Proietti, Garinei, Bozza, Moretti and Marconi 2021), which focuses on forecasting wholesale food prices.

The forecasting models used in this study fall into three broad categories: time series models optimized using cross-validation, regression trees, and machine learning approaches. In the following, we briefly discuss the set of competing forecasting models.[1]

3.1 Time series models

Time series models forecast by leveraging the temporal patterns in the data. Such models may feature model selection, such as greedy search on the set of possible specifications, regularization, or model averaging. Forecasting models can be optimized using cross-validation techniques, whereby the parameters are tuned on a training set, and performance is evaluated on a validation set. This helps to select the most accurate model for a time series.

3.1.1 ARIMA

The first model is the ubiquitous univariate ARIMA model, whose regular and seasonal parameters are selected to minimize the Bayesian Information Criterion (BIC). The BIC is a popular model selection criterion that balances the performance of a model and its complexity measured by the number of parameters in the model. It penalizes overfitting by preferring a simpler model to a more complex model, unless the residual sum of squares of the more complex model is much smaller. When picking a model, the model with the lowest BIC should be preferred (Schwarz 1978). The ARIMA model selection follows the two-step procedure discussed in Hyndman and Khandakar (2008). Since the growth rates of the targets pass the weak stationarity test (KPSS test), we expect the optimal ARIMA models to be ARMA models, i.e. they do not require regular or seasonal differencing.

Table 2 reports the optimal ARIMA models for the target series. The parameters p and q (P and Q) denote the number of regular (seasonal) AR and MA terms; d (D) denotes the order of the regular (seasonal) differencing operators. The results confirm the weak stationarity (d, D) and the absence of seasonality (P, Q) in the growth rates of the target variables.

As a univariate model, ARIMA is a useful benchmark for the information and predictive power embodied in the past values of the time series and possibly in the current and past values of an error term. Richer forecasting models that feature indicators as independent variables are

[1] All the models used in this study have either been implemented in R by the authors or have been estimated using existing implementations in R.

expected to outperform the ARIMA model, except perhaps at very short time horizons where the own momentum of the time series may obscure the signal contained in the indicators. The BIC tends to select parsimonious models. Despite being (asymptotically) consistent, it does not necessarily identify the true model in a finite sample. Using the BIC criterion for model selection optimizes fit but not forecasting performance. While ARIMA models can be optimized for forecasting performance using cross-validation, we reserve explicit forecasting optimization for the following set of models that extend the univariate framework of the ARIMA model.

Table 2: Optimal ARIMA models.

	p	d	q	P	D	Q
Plant	2	0	0	0	0	0
Animal	3	0	1	0	0	0
Misc.	1	0	1	0	0	0
Alcoholic	2	0	0	0	0	0
Non-alcoholic	3	0	2	0	0	0

3.1.2 Hansen model selection and combination

A natural extension of the ARIMA model includes indicators as independent variables and uses the information contained in their time variation to improve predictive power. With M indicators and L ($L \geq 0$) lagged indicators, where the contemporaneous time series have $L = 0$, the number of conceivable model specifications equals ML. An exhaustive search on the set of all combinations of indicators and their lagged values would thus require estimating 2^{ML} different specifications, including the univariate specification. The computational burden of sifting through all conceivable specifications with 317 indicators is prohibitive, especially if autoregressive terms in the target series or moving average terms are also to be included.

To keep the computational burden manageable, we filter the indicators based on the magnitude of their leading cross-correlations with the target series before including them in the model search. We thus hope to retain *promising indicators* and discard those indicators with low predictive power. The selection procedure then operates on the subset of *promising indicators*.

For a pair of time series x_t and y_t, the cross-correlation at time shift $h \in \mathbb{Z}$ is defined as:

$$\rho(h) = \frac{\mathrm{E}_h \left(x_{t-h} - \mu_x\right)\left(y_t - \mu_y\right)}{\sigma_x \sigma_y},$$

where μ_x, μ_y and σ_x, σ_y are the constant means and standard deviations. When the time shift is positive, $h > 0$, the correlation coefficient shows the degree of lead of the indicator series x_t on the target series y_t. We call x_y a *promising leading indicator* for the target series y_t at horizon h if the magnitude of the correlation is sufficiently high, $|\rho(h)| > \bar{\rho}$. The range of reasonable threshold $\bar{\rho}$ value for a cutoff criterion will depend on the sample and may also depend on the horizon h. We found that in our sample, $\bar{\rho} = 0.5$ proved to be a reasonable choice, retaining sufficiently many indicators at all horizons. The correlation threshold should be set in such a way that at least one of the indicators (ideally several indicators) exceeds it. This indicator would then compete with the optimal univariate model.

Our time series data suffers from a multicollinearity problem in that some indicators have almost identical growth rates and, as a consequence, almost identical correlations with the target series. To address this problem, we first sorted the indicators in descending order of correlation and then removed all indicators that differed from the previous indicator in the sorted list by less than ρ_δ in absolute value; in our case $\rho_\delta = 0.05$. In other words, we removed all the indicators that were within the tolerance of the previous indicator (i.e. too similar).[2]

To identify models that minimize the forecast error for a given forecast horizon, we use the model selection procedure proposed by Hansen (2008). The **Hansen Select** procedure is based on the fact that leave-one-out cross-validation (LOOCV) errors approximate the one-step-ahead forecast error of the ordinary least squared (OLS) regression. The data set is divided into n subsets, where n is the total number of observations. Each subset contains a single observation, and the model is trained n times, each time omitting a single observation as the test sample and using the remaining $n-1$ observations as the training sample. The model's performance is then evaluated on the left-out observation. It is important to emphasize that the one-step-ahead forecast error is the appropriate performance metric because we employ separate models for each forecast horizon when predicting the target series. Consequently, each forecast is essentially a one-step-ahead forecast of a specific model.

Let m be a model for a target $\mathbf{y}$ featuring a full-rank regression matrix containing some indicators $\mathbf{X}$. The model m is completely defined by the indicators included in $\mathbf{X}$.[3] The model includes a constant term so that the first column of $\mathbf{X}$ contains the unit vector. Let n be the number of observations, or the number of rows of $\mathbf{X}$.

Denote the OLS estimate of the model m by $\hat{\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{y}$ and let $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$ be the vector of residuals. Define a squared matrix $\mathbf{M} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}$. The matrix $\mathbf{M}$ is sometimes called the residual maker matrix, as $\mathbf{e} = \mathbf{y} - \mathbf{M}\mathbf{y}$. Let the vector $\mathbf{h}$ be the diagonal of $\mathbf{M}$, i.e. $\mathbf{h} \equiv diag(\mathbf{M})$. Then, the vector $\tilde{\mathbf{e}}$ with elements

$$\tilde{e}_i = \frac{e_i}{1-h_i} \text{ for all } i = 1, \ldots, n$$

approximates one-step-ahead forecast error of m in a LOOCV, and

$$CV = \frac{\tilde{\mathbf{e}}^\top\tilde{\mathbf{e}}}{n}$$

approximates the mean squared forecast error. Given a set of competing models $m = 1, \ldots, M$, the optimal model minimizes the CV criterion.

Model selection proceeds in two steps: first by selecting the lagged values of the target variables and then by selecting the indicators and their lagged values. The two-step procedure is applied separately to each forecasting horizon, yielding twelve optimal specifications that minimize the LOOCV criterion. Thus the number of competing models equals $2^{L(1+M)}$, where the first term stands for up to 2^L combinations of the lags of the target variable and the second term for all combinations of the indicators and their lagged values 2^{ML}.

There are many alternatives to the Hansen model selection procedure, including the classical AIC and BIC criteria, the Mallows criteria, and predictive least squares.[4] The advantage of the

[2]Formal methods for detecting multicollinearity include, e.g., inflation variance factors in Kutner, Nachtsheim and Neter (2004).

[3]Here and below, we omit the target and horizon to keep the notation simple, although each model is specifically optimized to forecast for a specific horizon.

[4]The literature on model selection and predictive combination is quite extensive, drawing on more than half

Hansen approach is its computational simplicity. It does not require estimating the model for each of the n potential training samples, where n is the number of observations. The root-mean-squared forecast error is computed directly from the model residual e_i at the validation observation and the weight h_i. The forecast errors are estimated directly from the fit errors (residuals) without the need to re-estimate the model each time a single observation is transferred from the training sample to the test sample. Another advantage of the Hansen approach is its robustness to heteroskedasticity of model disturbances.

Table 2 compares the number of *promising indicators* that meet the correlation criteria by forecast horizon, the optimal number of lags of the target, and the number of indicators or their lags selected by the Hansen procedures. The number of such indicators decreases as the horizon increases but this is to be expected as it is easier to find indicators that contain information about the near future than about the distant future. The Hansen selection tends to select parsimonious models. For this reason, we supplement the Hansen selection with the Hansen combination, which weights the forecasts of several models to improve forecast accuracy.

The **Hansen Combine** procedure seeks non-negative weights $\mathbf{w}$ for a set of competing models indexed by $m = 1, \ldots, M$ using a constrained quadratic program. Let $\mathbf{D} = [\mathbf{e}(1), \ldots, \mathbf{e}(M)]$ be a matrix of dimensions $n \times M$, where n is the number of observations and M the total number of models. The quadratic program is given by:

$$\arg\min_{\mathbf{w}} \quad \left\{ \frac{1}{2} \mathbf{w}^\top \mathbf{D} \mathbf{w} \right\}, \text{ such that } w_m \geq 0 \text{ for all } m = 1, \ldots, M \text{ and } \sum_{m=1}^{M} w_m = 1.$$

It features $M + 1$ constraints, of which there are M inequality constraints and a single equality constraint. Some kind of conditioning of the matrix $\mathbf{D}$ may be required when there are many observations and models to ensure that it is numerically positive semi-definite.[5]

The Hansen selection chooses the best model, whereas Hansen combination allows for the inclusion of multiple weighted models, potentially including more indicators into the final forecast. Selection is thus a special case of combination. The number of relevant indicators that meet the correlation criteria decreases with the forecast horizon (Table 2). The Hansen combination forecast includes all indicators, as each of them is present in some model with a nonzero weight.

The Hansen selection and combination methods are computationally intensive when the number of indicators is large. Hansen selection considers all combinations of indicators and their lags, and Hansen combination solves a quadratic program with inequality constraints in which the number of variables is given by the number of combinations. To apply the Hansen methods, we had to restrict the data set to those indicators that have a reasonable lead over the target at a given horizon. We will continue to use this smaller data set of *promising indicators* as a benchmark. Next we discuss models that can be estimated using the entire data set of 317 indicators and their lagged values.

a century of research, see Bates and Granger (1969), Akaike (1973), Mallows (1973), Granger and Ramanathan (1984), Hendry and Clements (2004), Stock and Watson (2006), Timmermann (2006), Hansen (2007).

[5]This problem often arises in the context of correlation matrices. We use the method proposed by Higham (2002). For an alternative method, see Knol and ten Berge (2006).

Table 3: Variable selection using correlations and Hansen methods.

Horizon	1	2	3	4	5	6	7	8	9	10	11	12
							Plant					
Lead correlation	8	7	7	7	6	6	5	5	4	4	4	3
Hansen Select: Target lags	1	2	2	2	2	2	3	3	3	3	2	2
Hansen Select: Indicators	5	1	4	3	5	3	4	4	3	2	4	2
Hansen Combine: Target lags	1	2	2	2	2	2	3	3	3	3	2	2
Hansen Combine: Indicators	8	7	7	7	6	6	5	5	4	4	4	3
							Animal					
Lead correlation	9	9	9	9	9	8	8	7	7	7	7	6
Hansen Select: Target lags	4	3	2	3	3	3	3	3	3	3	2	2
Hansen Select: Indicators	5	4	5	5	6	6	6	4	4	7	5	4
Hansen Combine: Target lags	4	3	2	3	3	3	3	3	3	3	2	2
Hansen Combine: Indicators	9	9	9	9	9	8	8	7	7	7	7	6
							Misc.					
Lead correlation	6	6	6	6	6	6	7	7	7	8	8	7
Hansen Select: Target lags	4	3	3	3	2	3	3	3	4	3	1	1
Hansen Select: Indicators	4	3	4	4	4	3	6	5	2	4	4	4
Hansen Combine: Target lags	4	3	3	3	2	3	3	3	4	3	1	1
Hansen Combine: Indicators	6	6	6	6	6	6	7	7	7	8	8	7
							Alcoholic					
Lead correlation	5	5	4	4	4	4	4	4	3	3	2	1
Hansen Select: Target lags	3	3	3	3	2	2	2	3	4	3	2	3
Hansen Select: Indicators	1	1	2	3	1	2	2	1	2	1	2	1
Hansen Combine: Target lags	3	3	3	3	2	2	2	3	4	3	2	3
Hansen Combine: Indicators	5	5	4	4	4	4	4	4	3	3	2	1
							Non-alcoholic					
Lead correlation	8	8	8	8	8	8	9	9	9	9	9	8
Hansen Select: Target lags	4	4	4	4	4	3	3	4	3	3	1	1
Hansen Select: Indicators	5	4	4	5	6	6	6	4	3	4	7	4
Hansen Combine: Target lags	4	4	4	4	4	3	3	4	3	3	1	1
Hansen Combine: Indicators	8	8	8	8	8	8	9	9	9	9	9	8

The table shows the number of indicators that meet the correlation criteria by forecast horizon, the number of lags of the target, and the number of indicators or their lags selected by Hansen model selection and model combination. The number of relevant indicators that meet the correlation criteria decreases with the forecast horizon. The Hansen selection tends to select a parsimonious model and is therefore complemented by the Hansen combination, which is more comprehensive and potentially exploits a broader set of information.

3.1.3 Regularized regressions

The correlation patterns revealed that many indicators are highly correlated. This is not surprising, as most of the indicators are producer, wholesale, or consumer prices that underlie similar inflation trends. High correlation causes the moment matrix $\mathbf{X}^\top\mathbf{X}$ used to compute the OLS

estimate to be near-singular. The presence of many highly correlated indicators motivates the use of regularization to improve the OLS estimate and achieve variable selection.

Regularization adds a penalty term to the loss function of a multiple-regression model with the aim of balancing between model complexity and out-of-sample performance. In the context of a bias-variance trade-off, regularization trades a small increase in bias for a substantial decrease in variance. Regularized regression offers superior performance compared to standard OLS regression in scenarios where multicollinearity is present and variable selection is required.

The lasso and ridge regressions impose $\mathcal{L}_1$ ($\|\cdot\|_1$) and $\mathcal{L}_2$ ($\|\cdot\|_2$) penalties on the OLS estimate. The models are optimized using 10-fold cross validation that partitions the data into 10 approximately equal-sized subsets, one of which is retained for testing.[6]

Elastic Net linearly combines the two penalties, as

$$\hat{\beta} \equiv \arg\min_{\beta} \quad \left\{ \|\mathbf{y} - \mathbf{X}\beta\|^2 + \lambda \left[0.5(1-\alpha)\|\beta\|_2^2 + \alpha\|\beta\|_1 \right] \right\}.$$

The ridge and lasso regressions emerge as special cases when $\alpha = 0$ and $\alpha = 1$, respectively. We estimate ridge and lasso regressions separately, as well as their combination with $\alpha = 0.5$. The optimal choice of the parameter λ is obtained using 10-fold cross validation.

Ridge regression augments the loss function with a penalty term equal to the square of the magnitude of the coefficients ($\mathcal{L}_2$ norm). The penalty term encourages the model to keep the regression coefficients small, effectively shrinking them towards zero. This reduces the impact of individual indicators without excluding them completely, allowing all indicators to contribute to the forecast. Ridge regression is particularly useful in the presence of multicollinearity among the indicators.

Lasso regression adds a penalty term equal to the absolute value of the coefficients ($\mathcal{L}_1$ norm). The penalty term encourages model parsimony by letting some coefficients to vanish. Lasso regression thus performs variable selection, which is particularly useful when dealing with many indicators, only a few of which are likely to be relevant. Preliminary correlation analysis has shown that this is indeed likely to be the case in our data. Last but not least, the more parsimonious the model, the easier it is to interpret.

3.2 Regression trees

Regression trees are a popular ensemble learning method used for both regression and classification. They are used most effectively when processing high-dimensional data with many indicators that may contain outliers and missing observations. Regression trees and Bayesian additive regression trees (BART) leverage the strengths of recursive decision trees to capture nonlinear relationships in the data and combine them with various averaging methods to reduce forecast variance and avoid overfitting. By using a Bayesian approach, BART is particularly useful for expressing forecast uncertainty. These and other qualities make both models valuable for analyzing and forecasting time series data.

Random forests average multiple regression trees with the aim of improving the accuracy and stability of forecasts. In this study, we use Breiman's (2001) random forest algorithm equipped with randomized node optimization and bootstrap aggregation (bagging). In random

[6]For a comprehensive treatment of regularized regressions and related concepts, see James, Witten, Hastie and Tibshirani (2013), Hastie, Tibshirani and Wainwright (2015), Efron and Hastie (2016). An example of an application that uses the implementation of ridge, lasso, and elastic net regressions in R can be found in Melkumova and Shatskikh (2017).

forests, forecasts are obtained from an average of regression trees, each of which relies on a random sample and random choice of indicators. Random forests are particularly useful for extracting non-linear patterns in noisy data. However, they struggle to capture long-term trends and may require additional techniques to address seasonality.[7]

Bayesian Additive Regression Tree (BART): The effectiveness of regression trees can be greatly improved by gradient boosting. Boosting combines weighted trees, with each new tree predicting the residual of the previous fit. Each new tree aims to capture the variations that were not captured by the existing trees in the ensemble. BART is closely related to random forests with bagging and gradient boosting, see Chipman, George and McCulloch (2010), or Hill, Linero and Murray (2020). Each new tree is constructed randomly as in bagging, and each tree attempts to capture variation not previously captured by the model as in boosting. The main difference lies in the way new trees are generated by perturbations based on partial residuals from the previous iteration. BART extends the idea of boosting by placing a prior distribution on the ensemble of regression trees, allowing for uncertainty estimation in the predictions. It does so within a Bayesian framework to derive the posterior distribution of the parameters using Markov Chain Monte Carlo sampling or variational inference.

The use of BART in time series forecasting offers several benefits. It can handle complex and non-linear relationships, capture interactions between predictors, and provide probabilistic forecasts. BART is particularly useful when dealing with time-varying dynamics, irregularly spaced data, and situations where there is limited domain knowledge or prior information.

3.3 Machine learning

Machine learning algorithms learn patterns and relationships in time series data. Some commonly used machine learning algorithms for time series forecasting include support vector machines and neural networks. Extreme Learning Machines (**ELM**) belong to the family of feed-forward neural networks. EMLs are known for their computational efficiency and fast learning rates. They offer an alternative to traditional neural networks that overcome the limitations of the latter in terms of training time and generalization performance.[8]

The ELM consists of a single hidden layer of neurons, randomly initialized with weights, and an output layer. During training, the input data is fed through the hidden layer, and the output weights are directly computed using a regression technique. The model in this study uses one of the four regression types in the output layer: the conventional linear regression as well as ridge, lasso and step regressions. The neurons in the hidden layer use a given activation function, here a sigmoid function, to transform the input data into a higher-dimensional feature space. The output weights are then obtained by solving a linear system of equations, which makes the training process exceptionally fast.

ELMs have gained popularity for time series forecasting due to their ability to handle large amounts of data efficiently. The random initialization of the hidden layer weights enables ELMs to capture complex temporal patterns and nonlinear relationships in time series data. The algorithm requires preprocessed time series, for example data normalization. ELMs have been applied to a wide range of time series forecasting problems. Their simplicity and fast training process make them suitable for handling large-scale data sets and real-time applications.

[7] More information on regression trees can be found in Chipman, George and McCulloch (2006), Hastie, Tibshirani and Friedman (2009), Sheppard (2017).

[8] See, Ding, Zhao, Zhang, Xu and Nie (2013), Huang, Zhu and Siew (2004), Huang, Zhu and Siew (2006).

Both multilayer perceptron (**MLP**) and ELM are supervised machine learning models used for classification and regression tasks. However, there are some key differences between the two. The MLP uses backpropagation to adjust the weights of the network during training. In contrast, ELM uses a single-pass training algorithm where the weights are randomly generated and only the weights in the output layer are adjusted during training. The MLP has multiple hidden layers between the input and output layers, whereas ELM has only one hidden layer. The MLP is prone to overfitting, especially when the number of hidden layers is large, while ELM is less prone to overfitting due to its randomly generated weights. The MLP requires a significant amount of computation to adjust the weights of multiple hidden layers during training, whereas ELM has a simpler training process that requires less computation. In summary, MLP is a more complex model with multiple hidden layers and a backpropagation training algorithm, while ELM is a simpler model with a single hidden layer and a single-pass training algorithm. ELM has advantages in terms of computational complexity and is less prone to overfitting but may not perform as well as MLP in certain applications.

4 Forecasting performance

Let us next turn to the forecasting performance of the best models or the average forecasting performance of all models at different episodes and horizons. The test episodes labeled 2018, 2019, 2020, 2021 refer to the year in which the forecasts are made. We do not evaluate forecasting performance prior to 2018, since withholding data for out-of-sample evaluations would leave the training set that begins in January 2001 too short for some models. Starting in 2018 allows us to retain at least 200 monthly observations for fitting the models. The four test episodes cover the recent past. Excluding 2022 from the evaluation allows us to retain twelve months as the maximum forecasting horizon for each episode.

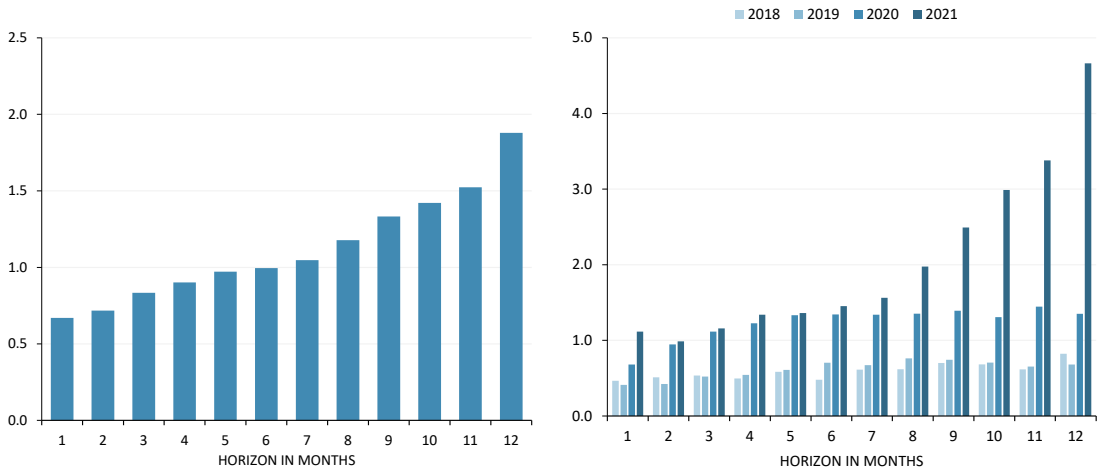
In each month of a given test episode, twelve out-of-sample forecasts for the following twelve months are computed with a set of models estimated on the data available up to that time. The forecasts cover the current and following year, with the first one being a one-month-ahead forecast of January 2018 for February 2018 and the last one being the twelve-months-ahead forecast of December 2021 for December 2022. The statistics reported for a particular year should be taken either as the most accurate (best) forecast or the average of all forecasts that would have been computed in that year. The averages for a given forecasting horizon aggregate all one-month-ahead forecasts, all two-months-ahead forecasts, etc.

The recent years have been hallmarked by an exceptionally dynamic inflationary environment. The choice of test episodes is also motivated by the desire to compare forecasting performance before and after the severe economic downturn caused by the pandemic and the subsequent rise in inflation. We thus compare the calmer times with the times of crises. We expect forecast errors to be smallest for the 2018 forecasts, as inflation was moderate in 2018 and 2019, and largest for the longer-term 2021 forecasts when inflation soared following the outbreak of the Ukraine war in February 2022. We also expect that performance and the usefulness of models will gradually return to its pre-crisis levels once the current inflation dynamics abate.

There exist many quantitative measures of forecasting accuracy, the most common ones being the root-mean-squared error (RMSE) and the mean absolute error (MAE). Both measures carry the same units as the target. This makes it convenient to compare errors with the mean of the target and the standard deviation. The mean conveys the magnitude of the target time series,

and the standard deviation conveys the amplitude of its variation.[9] More volatile series should be more difficult to forecast, especially if the volatility emanates for random idiosyncratic shocks rather than the current dynamics of the leading indicators acting as exogenous predictors. We express the forecasting accuracy of a model using the MAE, because it is less sensitive to outliers than the RMSE.

Figure 3: Best model's forecast MAE by horizon



(a) Average over all targets and forecast years

(b) Average over all targets

The mean absolute error (MAE) in percentage points tends to increase with the forecast horizon for all years and has increased dramatically since 2020.

Before comparing the forecasting performance of individual models, it is useful to get a broad view on the forecasting accuracy they provide. Figure 3 shows the MAE of most accurate (best) models averaged over all targets and test episodes. The best models are listed in Figure B.1 of the Appendix B. A ranking of the models will be discussed in detail at the end of this section. Here we pause to remark on the expected steady decline in the forecast accuracy with the horizon in the left panel of Figure 3. One-year-ahead forecasts are roughly three times less accurate than one-month-ahead forecasts, the error increasing with the horizon at a faster than linear rate. The right panel of Figure 3 presents the errors separately for the four test episodes. The errors have increased substantially during 2020 and 2021. The increase is disproportionately high at longer forecast horizons, but even one-month-ahead errors have roughly doubled. This is especially true for the forecasts made in 2021, which cover 2022 in the longer term. In February 2022, the war between Russia and Ukraine has triggered a major global shock in energy prices. This shock amplified the inflation momentum gained during the recovery from the pandemic in 2021, leading to unprecedented inflation rates in the second half of 2022.

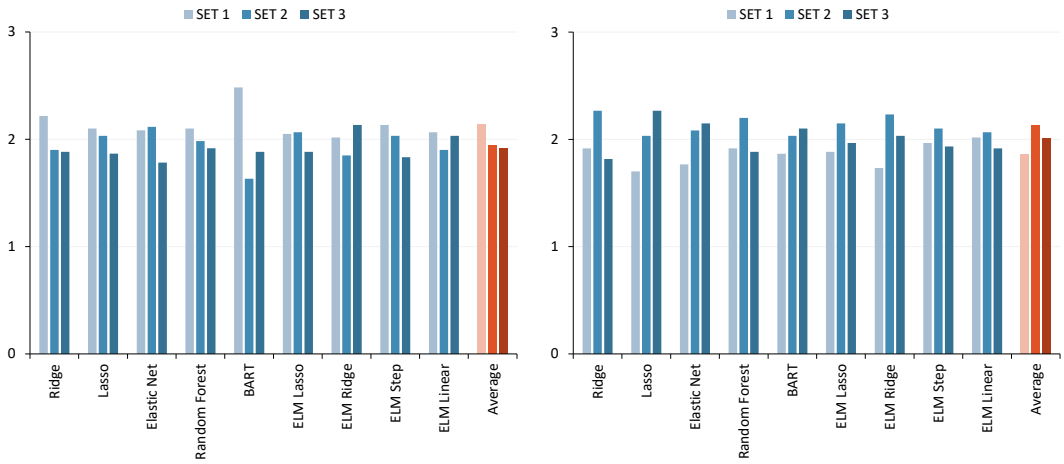
The time episode for out-of-sample forecasts is not the only element of the design of a data set used in evaluating the models. With a maximum of 6 lags and 317 indicators, the total number of variables that can enter a model equals $6 + 7 \cdot 317 = 2225$. To keep the size of the

[9]This information can be combined using a coefficient of variation defined as the ratio of the standard deviation to the mean. The average inflation rates in our sample are never too low, so that the coefficient of variation is not overly sensitive to small changes in the mean.

data sets to a manageable size, we limit ourselves to the following three designs (Set 1, 2, 3) for each episode and horizon.

Set 1 contains the *promising indicators* and their lags. These indicators meet the correlation criteria. In addition this set contains the lags of the target that have been identified using the first stage of the Hansen selection procedure. Recall that the Hansen selection procedure first picks the lags of the target and then picks indicators and their lags from the set of *promising indicators*. We noted in Section 3.1.2 that the computational complexity of the Hansen selection (greedy search) and combination (quadratic program) methods means that they can only be applied to a subset of *promising indicators*, which is much smaller than the set of all indicators. Set 1 is the smallest of the three. The number of variables in the first set is given by the figures for Hansen select shown in Table 3. Set 2 includes the own lags of the target that have been identified using the two-stage Hansen model selection procedure and all contemporaneous indicators. Set 3 features all own lags of the target and all contemporaneous indicators. Sets 2 and 3 thus feed all indicators to the forecasting models, except their lagged values. The models then rely on their selection and aggregation mechanism to produce optimal forecasts. Set 3 contains 347 variables, i.e. six lagged values of the target and all 317 contemporaneous indicators.

Figure 4: Relative ranks of models by set



(a) Forecast year 2018

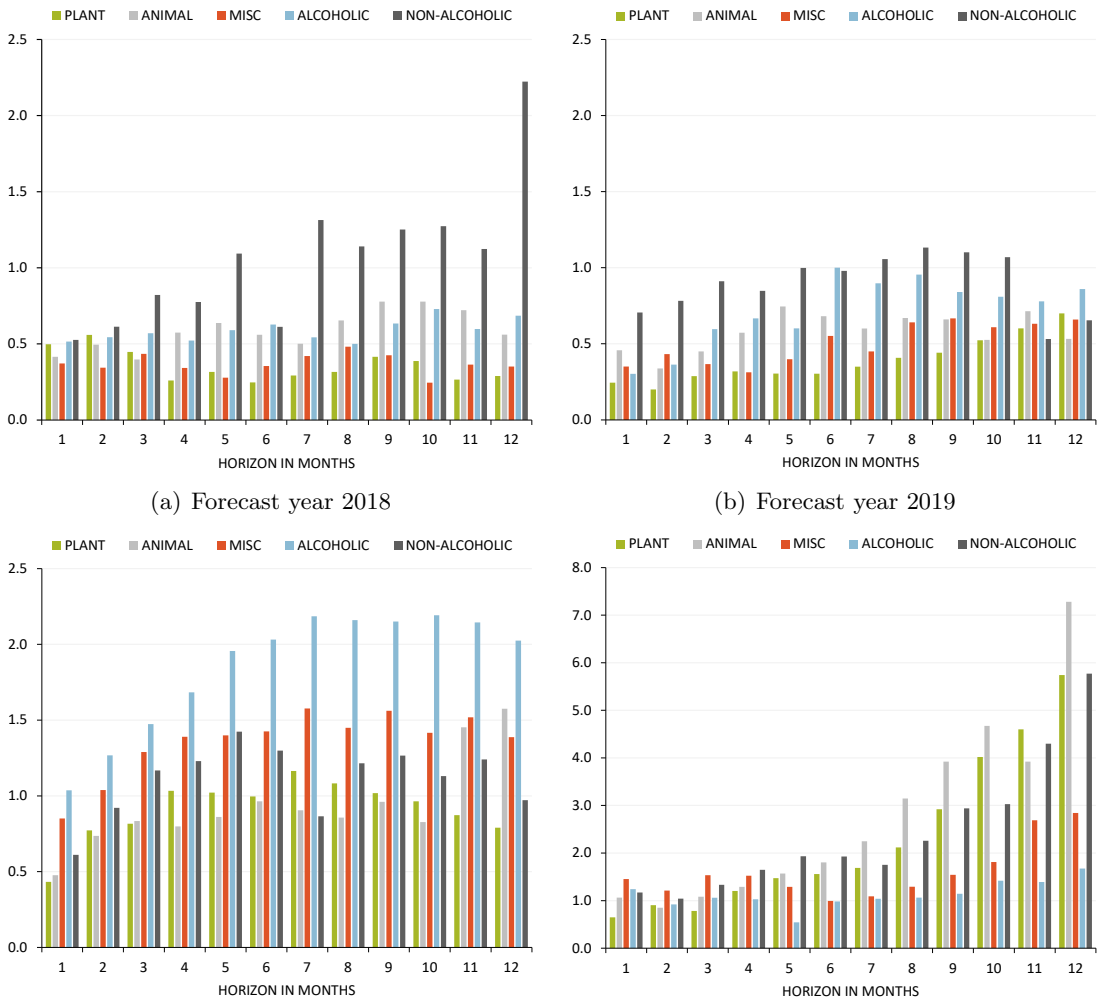
(b) Forecast year 2021

Set 1 contains the *promising indicators*. Set 2 includes the target lags selected by the Hansen procedure and all contemporaneous indicators. Set 3 includes all target lags and all contemporaneous indicators. The first and most parsimonious set tends to produce least accurate forecasts in 2018 and most accurate forecasts in 2021.

Figure 4 shows the relative ranks of the models between the three sets for the 2018 and 2021 forecasts. For each model that can be estimated with all three sets, the version with the lowest MAE is given rank one and the version with the highest MAE is given rank three. Relative ranks are calculated separately by test episode and horizon. Relative ranks (blue bars) were averaged across all targets and horizons. The red bars show the average of the ranks of all models.

The forecasting performance of the models varies across the sets. While the forecast errors for 2018 show the expected pattern that larger and presumably richer data sets lead to smaller forecast errors, this does not appear to be the case in 2021, where Set 1, based on the *promising indicators*, performs best. Note that the Hansen methods are not included in this figure because they were computed using only Set 1.

Figure 5: Best model's forecast MAE by target and horizon



(a) Forecast year 2018

(b) Forecast year 2019

(c) Forecast year 2020

(d) Forecast year 2021

The forecast errors of best models for beverages tend to be higher than those for food products, especially for non-alcoholic beverages in 2018 and for alcoholic beverages in 2020. The forecast errors for plant-based and animal-based foods increased dramatically in 2021.

Figure 5 shows the forecast errors (MAE) separately by targets, episodes and horizons. The forecast errors for beverages are higher than for the prices of food. This can be seen from the elevated forecast errors for non-alcoholic beverages in 2018 (dark gray), all beverages in 2019 (blue and dark gray) and alcoholic beverages in 2020 (blue). It is therefore quite remarkable that the forecast errors for beverages are relatively low during 2021, i.e., during the period in which their prices have risen from an average of 1 percent for forecasts made in 2018-2020 to more than 4 percent for forecasts made in 2021. This combination of low predictive power during calmer periods and high predictive power during periods of extreme inflation suggests that the available indicators are not particularly informative. The consequence is that the recent price dynamics of a target remains a reasonable predictor of its future dynamics. This is indeed the case, as Table 3 shows that for alcoholic beverages only few indicators meet the correlation criterion of Section 3.1.2. The list of best models in Figure B.1 shows that an optimally selected ARMA model often emerges as the best model for beverages in 2018 (alcoholic beverages at horizons 7-12) and 2019 (non-alcoholic beverages at horizons 1-5). The relatively good performance of the univariate ARMA model confirms the lack of informative indicators for alcoholic beverages.

Table 4 at the end of this section summarizes the forecast errors (MAE) for windows of two consecutive years, indicated by year in which the forecast is computed, supplemented by basic summary statistics for the targets. The forecast errors of the best (most accurate) models at different horizons are averaged by trimesters of 1-4 months, 5-8 months, and 9-12 months. Taking the inflation rates for animal, plant, and miscellaneous food items together, forecast errors have been around 0.8 to 1.7 percentage points, which is roughly within 22 to 45 percent of the standard deviation and about 30 to 60 percent of the mean. The forecasts for beverages have been slightly less accurate with average errors of around 1.1 to 1.8 percentage points, or roughly within 33 to 55 percent of the standard deviation and 50 to 80 percent of the mean. Comparing the error of the forecasts made in 2018 and 2021, the errors of food price forecasts have increased by roughly 0.8 percentage points for horizons of up to three months to 3.7 percentage points on the longer end of the twelve-month horizon. The forecast errors for beverage prices have also risen but only by 2.0 percentage points on the longer end.

The final aspect we investigate is the relative forecasting performance of the models. Figure B.2 in the Appendix summarizes the distribution of forecast errors by target and horizon for the 2018 and 2021 forecasts. The 2018 forecasts show a large spread in the forecast errors for alcoholic beverages and for plant-based foods forecast 7 to 10 months ahead. For the beverages, we see an increase in the spread of forecast errors with the forecast horizon. The spread of forecasts for 2021 is much clearer, with higher median errors but tighter distributions around the median. This reflects the fact that during the crisis the performance of all models deteriorates, so the relative performance of the models stays comparable, resulting in fewer exceptionally good and exceptionally bad models in the group.

Figure B.1 shows the most accurate model by target, episode, and horizon. The color gradient is scaled from green (low MAE) to red (high MAE), relative to all errors, i.e. taken together over all cells representing targets, episodes and horizons. The predominance of red at the bottom shows the decrease in forecast accuracy toward the end of episode 2021, except for alcoholic beverages, where the largest errors occurred in episode 2020. The lowest errors occur in the forecasts for plant-based food products and miscellaneous food products during the 2018 episode. To get a better sense of the relative forecasting performance, we rank the models using weighted rankings r_{short} and r_{long}. The weighted ranks reflect the relative performance at short and long horizons. We compare the weighted ranks to its unweighted counterpart r_{equal}.

Let r be the (raw) rank of a model, such that $r = 1$ for the best model and $r = M$ for the

worst model, where each rank is computed for a given target, episode, and horizon. Figure B.1 shows all models for which $r = 1$ holds. The horizon-weighted cumulative ranks are defined as

$$r_{short} = \sum_{h=1}^{H} \frac{(H-h+1)}{\bar{h}} r, \quad r_{equal} = \sum_{h=1}^{H} r, \quad r_{long} = \sum_{h=1}^{H} \frac{h}{\bar{h}} r.$$

were H denotes the farthest forecast horizon and $\bar{h}$ is the average of the numbers $1, \ldots, H$ (here $H = 12$ and $\bar{h} = 6.5$). The normalization by the average rank makes the three ranks comparable.

The first weighted rank r_{short} assigns progressively smaller weights to increasing forecasting horizons. This simple linear weighting scheme with descending weights emphasizes the short-term forecasting performance of the model. Conversely, the second weighted rank r_{long} uses ascending weights to emphasize the long-term performance. The unweighted rank treats all forecast horizons equally and thus serves as a benchmark for the weighted ranks. The weighted rankings help finding the model that consistently outperforms other models for different time horizons. However, if one model performs best (low rank) in the short-term (1-4 months) and worse than other models (relative higher rank) in long-term (9-12 months), using a combination of several models may be superior to using a single model.

Figures B.3 to B.7 summarize the cumulative horizon-weighted ranking of the models for the 2018 and 2021 forecasts. For plant-based food products, we see that machine learning methods (ELM) tend come on top of other models in 2018. The horizon-weighted rankings for these models tend to be better (lower) for the long term then for the short term. The comparative performance of the univariate ARMA model deteriorates from 2018 to 2021, a common pattern as we shall see below.

For animal-based food products, we see good relative performance of time series models and regression trees, with the ridge regression and BART showing the lowest forecast errors for this target for all three data sets. The ARMA model performs surprisingly good. This pattern holds for 2021 as well, apart from the ARMA model, whose relative performance deteriorates drastically in times of crisis, especially over long forecasting horizons. The machine learning models for this target tend to have the worst performance of all competing models.

In the case of miscellaneous food products, we remark on the relatively good performance of the ARMA model, the ridge regression, and the random forest. The quality of the ARMA model deteriorates sharply between 2018 and 2021.

For alcoholic beverages, the time series models (Hansen methods and regularized regressions) as well as random forest consistently rank higher in the 2018 episode, but only when estimated using the smallest data set that contains the *promising indicators*. Essentially, this implies that the smaller the number of indicators the better the forecasting performance, strongly suggesting the absence of informative indicators for this target. The fact that the univariate ARMA model ranks highest supports this conjecture. The story changes in 2021, where the richer data sets (Set 2, 3) deliver relatively better models, adding BART to the set of good models. The ridge regression tends to perform best among all models, whereas the performance of other time series models deteriorates, especially when they are estimated using richer data sets.

In the set of forecasts of non-alcoholic beverages computed in 2018, random forest performs best by a significant margin, followed by ARMA and BART. The machine learning models perform worse than other types of models in all episodes. The ARMA model comes out worst in 2021. The remaining time-series models rank among the worse in this most recent episode. The tree-based models rank relatively better especially in the long-term, which is promising given the general difficulty of making long-term forecasts during a crisis.

The above results show no overall winner in terms of forecasting accuracy for all targets, episodes, and time horizons. Within the best models listed in Figure B.1, time series and regression tree models tend to outperform machine learning models for medium-term (5-8 months) to long-term (9-12 months) forecasts, which is broadly true for both calmer periods (2018) and crisis periods (2020, 2021). Time series models perform quite well, especially lasso regression and the more general elastic net. The lasso regression tends to impose stricter model selection, which appears to be an advantage in our sample. The random forest model outperforms BART, except for the forecasts for the year 2021. Among the three types of models, machine learning models tend to perform worse, except for the forecasts for plant food in 2018 and 2019, miscellaneous items in 2020, and alcoholic beverages in 2019. Among the four ELM models, the models with linear and ridge functions in the output layer perform better than the lasso and step regression variants. Based on this assessment, if we had to choose a winner, we would vote for random forest as the best overall model, followed by BART and lasso regression.

Table 4: MAE of best models by trimester.

	Sample		MAE per trimester		
	Mean	Sd	h1-h4	h5-h8	h9-h12
Plant					
2018	0.8	0.6	0.6	0.4	0.4
2019	1.3	0.8	0.3	0.4	0.8
2020	1.9	0.9	1.0	1.2	0.9
2021	5.7	4.4	1.1	2.8	4.6
2018-2021	3.0	3.6	0.8	1.2	1.7
Animal					
2018	1.9	1.0	0.7	0.6	0.9
2019	2.2	0.9	0.6	0.8	0.9
2020	1.7	1.5	0.8	1.0	1.4
2021	6.7	7.5	1.1	2.5	5.6
2018-2021	4.1	5.3	0.8	1.2	2.2
Misc.					
2018	1.2	0.5	0.4	0.4	0.5
2019	1.3	0.9	0.4	0.5	0.7
2020	0.5	1.6	1.2	1.5	1.7
2021	2.7	3.7	1.7	1.4	2.5
2018-2021	1.8	2.5	0.9	1.0	1.4
Alcoholic					
2018	0.9	0.4	0.7	0.6	0.7
2019	1.2	1.3	0.7	1.0	0.9
2020	0.9	2.1	1.4	2.1	2.2
2021	2.3	2.6	1.3	1.1	1.7
2018-2021	1.6	1.9	1.0	1.2	1.4
Non-alcoholic					
2018	0.7	1.2	0.8	1.2	1.9
2019	0.8	1.2	1.0	1.3	1.0
2020	1.2	1.2	1.1	1.5	1.3
2021	6.1	6.1	1.7	2.0	4.8
2018-2021	3.1	4.7	1.2	1.5	2.3

The table shows the means and standard deviations of the yoy growth rates by target and trimester as well as the mean absolute forecast error (MAE) of the best model. Forecast errors increase with the forecast horizon in all episodes. They have increased steadily since 2018 and reached their maximum during the recent energy price crisis.

5 Current forecast

The above evaluation shows no clear winner among the models in terms of forecasting performance. This raises the question of how to derive the final forecast from individual forecasts. Splicing the forecasts of the best models for each horizon produces excessively volatile (jagged) future paths that are difficult to interpret. Smoother forecasts can be obtained by combining all forecasts to a final forecast for each horizon, which has the benefit of conveying the uncertainty of the models through the variation in the forecasts. This approach will be followed to obtain the current outlook for the inflation rates of the five product categories.

Forecasts can be combined in uninformative ways by using the median of all forecasts, supplemented with an interquartile range to express model uncertainty. This approach provides a final set of forecasts that is robust to outliers. The forecasts can also be combined in an informative way, for example using a weighted average of all individual forecasts, where the non-negative weights are inversely related to the average forecast error made in the past. Here we use the MAE of the latest forecasts computed in 2021.

Let f_m, $m = 1, \ldots, M$, be the forecasts of individual models for a given target and horizon, and e_m be their forecast MAE. The final (MAE-weighted) forecast is computed as:

$$f = \frac{\sum_{m=1}^{M} f_m/e_m}{\sum_{m=1}^{M} 1/e_m}.$$

The final error reduces to a simple average of individual errors when $e_m = e$ for all m.

Figure 6 shows final forecasts combined from individual forecasts using either the median (blue) or the inverse error weighting (red), as well as the interquartile range and the minimum and maximum forecasts. The broad picture is that inflation rates for all types of food products will gradually decline by the end of the forecast horizon in May 2024. The inflation rates remain positive, so that prices will continue to rise throughout 2024 (compared to the previous year). While some models predict negative inflation rates as early as the second half of 2023, the interquartile range remains clearly above zero for all targets and horizons. The median and error-weighted final forecasts align closely, except for animal-based food products, where the median remains consistently below the error-weighted forecast for all forecast horizons.

In terms of the implied future inflation dynamics, the forecasts anticipate that inflation rates in the second half of 2023 will be approximately $3\frac{1}{2}$ percentage points lower than those recorded in the first half of the year. The largest decrease in inflation rates is expected for animal-based food products, with a decline of $5\frac{3}{4}$ percentage points, while the smallest decrease is projected for non-alcoholic beverages, at $1\frac{3}{4}$ percentage points. In total for 2023, the animal-based food products would have the highest average inflation rate of $13\frac{1}{4}$ percent yoy, while non-alcoholic beverages are expected to have the lowest average inflation rate of $8\frac{1}{2}$ percent yoy. The second most dynamic category of products includes plant-based food products and alcoholic beverages. Their inflation rates for 2023 clock in at $12\frac{1}{2}$ percent and $11\frac{1}{4}$ percent yoy, respectively.

Looking ahead to the first half of 2024, a further decrease in the average inflation rate of approximately $4\frac{3}{4}$ percentage points can be expected across all five product categories compared to the average of 2023. The highest decreases in inflation rates are projected for animal-based food products, with a decline of $6\frac{1}{2}$ percentage points, while the lowest decreases are anticipated for non-alcoholic beverages, at $2\frac{3}{4}$ percentage points.

6 Concluding remarks

The current high inflation rates pose uncertainties about the persistence of inflation. The study aims to provide reliable forecasts for the following categories of products: plant-based food products, animal-based food products, alcoholic beverages, non-alcoholic beverages, and a residual category. The five categories are defined using weighted baskets of highly disaggregated CPI data.

This study advocates a multi-model paradigm that combines the strengths of a diverse set of forecasting models based on a rich set of monthly indicators, with separate models estimated for each forecast horizon (up to twelve months). The set of indicators includes monthly business cycle and trade indicators, prices of internationally traded commodities, import prices and exchange rates, producer prices and wholesale prices, wages, financial market and business sentiment indicators, and additional CPI components such as carbon fuels and electricity.

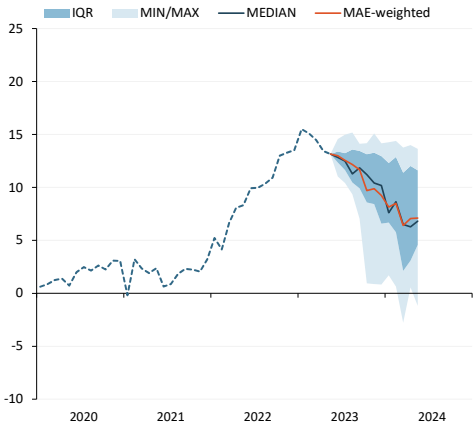
We evaluate the forecasting accuracy of time series models, regression trees, and machine learning models in the pre-pandemic and post-pandemic periods. We find that time series models (except the univariate ARMA model) and regression trees tend to work best, especially when the number of informative indicators is limited. This is particularly true at medium to long horizons. Since no individual model emerges as the definitive winner, the ensemble approach allows for a more robust and comprehensive analysis of future trends.

The forecasting performance of models has declined since the outbreak of the pandemic and subsequent political calamities that led to a sharp increase in energy prices and inflation rates. The multi-model paradigm leverages model diversity. We found that splicing the forecasts of the best models for each horizon results in erratic forecasts, while combining the forecasts of all models for each horizon results in smoother future paths that can be used to express the inherent model uncertainty. Beverages appear to be more difficult to forecast than foods, which suggests a lack of informative leading indicators of beverage price inflation in our data set.

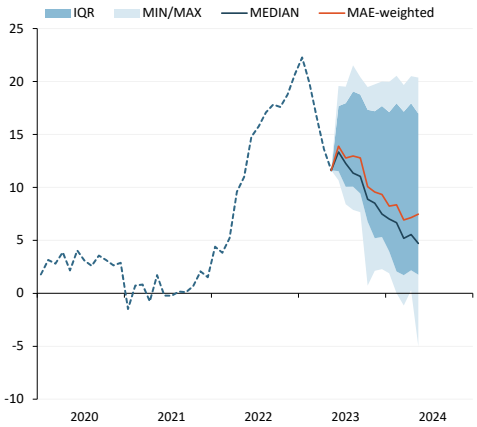
Based on weighted forecasts of all models (weighted by their recent errors), the current outlook suggests that elevated inflation rates for food products will persist in the foreseeable future in Austria. The analysis indicates that no deflationary trends can be expected until well into 2024. These findings underscore the challenges faced by policymakers and economic practitioners in formulating effective strategies to mitigate inflationary pressures and maintain price stability. The insights provided in this study contribute to the broader understanding of inflation dynamics and support policymakers in making informed decisions regarding monetary policy, fiscal measures, and competition regulation in the retail markets.

Acknowledgments: We would like to thank Peter Reschenhofer, Astrid Czaloun, Ursula Glauninger, Christine Kaufmann and Dietmar Weinberger for their help in preparing the data and editing the manuscript.

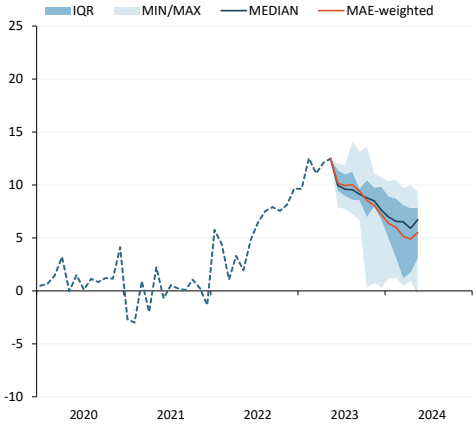
Figure 6: The current forecast till May 2024



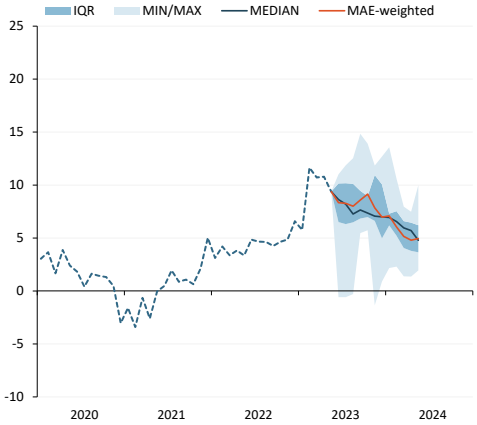
(a) Plant



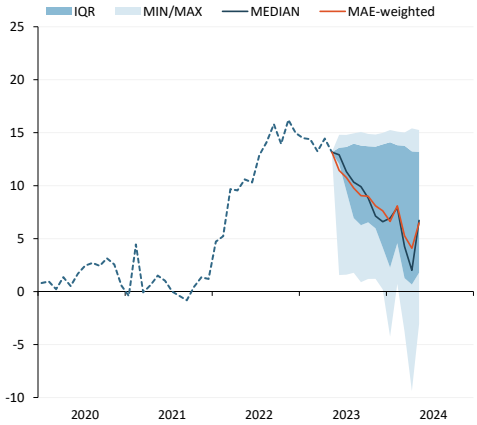
(b) Animal



(c) Misc.



(d) Alcoholic



(e) Non-Alcoholic

The figure shows the final forecasts combined from the forecasts of the individual models using the median (blue), or the inverse MAE weighting (red). The interquartile range and the minimum and maximum of the individual forecasts convey the uncertainty associated with the use of different forecast models.

References

Ahumada, H. and Cornejo, M.: 2016, Forecasting Food Prices: The Case of Corn, Soybeans and Wheat, *International Journal of Forecasting* **32**(3), 838–848.

Akaike, H.: 1973, Information Theory and an Extension of the Maximum Likelihood Principle, *in* B. N. Petrov and F. Csáki (eds), *Proceedings of the 2nd International Symposium on Information Theory*, Akadémiai Kiado, pp. 267–281.

Aparicio, D. and Bertolotto, M. I.: 2020, Forecasting Inflation with Online Prices, *International Journal of Forecasting* **36**(2), 232–247.

Bates, J. and Granger, C. W. J.: 1969, The Combination of Forecasts, *Journal of the Operational Research Society* **20**(4), 451–468.

Baumgartner, J.: 2022, WIFO Inflation Forecast 2022-2024 from December 2022. WIFO Research Briefs (25).

Baumgartner, J. and Sinabell, F.: 2021, Views on the Current Situation of Food Prices and their Expected Development. WIFO Research Briefs (10).

Breiman, L.: 2001, Random Forests, *Machine Learning* **45**, 5–32.

Cavallo, A.: 2018, Scraped Data and Sticky Prices, *The Review of Economics and Statistics* **100**(1), 105–119.

Cavallo, A. and Rigobon, R.: 2016, The Billion Prices Project: Using Online Prices for Measurement and Research, *Journal of Economic Perspectives* **30**(2), 151–78.

Chipman, H. A., George, E. I. and McCulloch, R. E.: 2006, Bayesian Ensemble Learning, *in* B. Schölkopf, J. Platt and T. Hoffman (eds), *Advances in Neural Information Processing Systems 19*, MIT Press, pp. 265–272.

Chipman, H. A., George, E. I. and McCulloch, R. E.: 2010, BART: Bayesian Additive Regression Trees, *The Annals of Applied Statistics* **4**(1), 266–298.

Ding, S., Zhao, H., Zhang, Y., Xu, X. and Nie, R.: 2013, Extreme Learning Machine: Algorithm, Theory and Applications, *Artificial Intelligence Review* **44**, 103–115.

Efron, B. and Hastie, T.: 2016, *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*, Institute of Mathematical Statistics Monographs, Cambridge University Press.

Faust, J. and Wright, J. H.: 2013, Forecasting Inflation, *in* E. Graham and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 2 of *Handbook of Economic Forecasting*, Elsevier, chapter 1, pp. 2–56.

for Reconstruction, I. B. and (IBRD, D.: 2023, Commodity Markets Outlook: Lower Prices, Little Relief.

Gorodnichenko, Y. and Talavera, O.: 2017, Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration, *The American Economic Review* **107**(1), 249–282.

Granger, C. W. J. and Ramanathan, R.: 1984, Improved Methods of Combining Forecasts, *Journal of Forecasting* **3**(2), 197–204.

Hansen, B. E.: 2007, Least Squares Model Averaging, *Econometrica* **75**(4), 1175–1189.

Hansen, B. E.: 2008, Least-squares Forecast Averaging, *Journal of Econometrics* **146**(2), 342–350.

Hastie, T., Tibshirani, R. and Friedman, J.: 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, 2 edn, Springer.

Hastie, T., Tibshirani, R. and Wainwright, M.: 2015, *Statistical Learning with Sparsity: The Lasso and Generalizations*, 1 edn, Chapman and Hall.

Hendry, D. F. and Clements, M. P.: 2004, Pooling of Forecasts, *The Econometrics Journal* **7**(1), 1–31.

Higham, N. J.: 2002, Computing the Nearest Correlation Matrix - A Problem from Finance, *IMA Journal of Numerical Analysis* **22**(3), 329–343.

Hill, J., Linero, A. and Murray, J.: 2020, Bayesian Additive Regression Trees: A Review and Look Forward, *Annual Review of Statistics and Its Application* **7**(1), 251–278.

Huang, G., Zhu, Q. and Siew, C.: 2004, Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks, *IEEE International Conference on Neural Networks - Conference Proceedings*, Vol. 2, IEEE, pp. 985–990.

Huang, G., Zhu, Q. and Siew, C.: 2006, Extreme Learning Machine: Theory and Applications, *Neurocomputing* **70**(1-3), 489–501.

Hyndman, R. J. and Khandakar, Y.: 2008, Automatic Time Series Forecasting: The forecast Package for R, *Journal of Statistical Software* **27**(3), 1–22.

James, G., Witten, D., Hastie, T. and Tibshirani, R.: 2013, *An Introduction to Statistical Learning with Applications in R*, Springer Texts in Statistics, 1 edn, Springer.

Jo, J. and Lusk, J. L.: 2016, Predicting Food Prices using Data from Consumer Surveys and Search, **230028**.

Knol, D. L. and ten Berge, J. M. F.: 2006, Least-squares Approximation of an Improper Correlation Matrix by a Proper One, *Psychometrika* **54**, 53–61.

Kuhns, A., Leibtag, E., Volpe, R. and Roeger, E.: 2015, How USDA Forecasts Retail Food Price Inflation, **1940**.

Kutner, M. H., Nachtsheim, C. J. and Neter, J.: 2004, *Applied Linear Regression Models*, 4 edn, McGraw Hill/Irwin.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y.: 1992, Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root, *Journal of Econometrics* **54**(1), 159–178.

Macias, P., Stelmasiak, D. and Szafranek, K.: 2023, Nowcasting Food Inflation with a Massive Amount of Online Prices, *International Journal of Forecasting* **39**(2), 809–826.

Mallows, C. L.: 1973, Some Comments on C_P, *Technometrics* **15**(4), 661–675.

Meade, B. and Rosen, S.: 1996, Income and Diet Differences Greatly Affect Food Spending Around the Globe, *Food Review/ National Food Review* **19**(3), 39–44.

Melkumova, L. E. and Shatskikh, S. Y.: 2017, Comparing Ridge and LASSO Estimators for Data Analysis, *Procedia Engineering* **201**, 746–755.

Menculini, L., Marini, A., Proietti, M., Garinei, A., Bozza, A., Moretti, C. and Marconi, M.: 2021, Comparing Prophet and Deep Learning to ARIMA in Forecasting Wholesale Food Prices, *Forecasting* **3**(3), 1–19.

Modugno, M.: 2013, Now-casting Inflation Using High Frequency Data, *International Journal of Forecasting* **29**(4), 664–675.

Monteforte, L. and Moretti, G.: 2013, Real-Time Forecasts of Inflation: The Role of Financial Variables, *Journal of Forecasting* **32**(1), 51–61.

Schwarz, G.: 1978, Estimating the Dimension of a Model, *The Annals of Statistics* **6**(2), 461–464.

Sheppard, C.: 2017, *Tree-based Machine Learning Algorithms: Decision Trees, Random Forests, and Boosting*, CreateSpace Independent Publishing Platform.

Stock, J. H. and Watson, M. W.: 1999, Forecasting Inflation, *Journal of Monetary Economics* **44**(2), 293–335.

Stock, J. H. and Watson, M. W.: 2006, Forecasting with Many Predictors, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, chapter 10, pp. 515–554.

Timmermann, A.: 2006, Forecast Combinations, *in* G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, chapter 4, pp. 135–196.

Whitmore Schanzenbach, D., Nunn, R., Bauer, L. and Mumford, M.: 2016, Where Does All the Money Go: Shifts in Household Spending over the Past 30 Years. The Hamilton Project.

A Target price series

Table 5: Plant-based foods (weights in percent).

Code	Name	Basket weight 2022	Basket weight 2023	Target weight 2022	Target weight 2023
002300	Nut cake	0.24	0.23	5.03	4.85
001000	Scone, handmade	0.22	0.22	4.60	4.68
000600	Rye bread	0.22	0.22	4.61	4.67
011400	Tomatoes	0.19	0.21	4.11	4.42
009700	Pickled cucumbers	0.18	0.17	3.75	3.56
000800	White bread	0.15	0.16	3.25	3.39
000900	Bread roll	0.14	0.15	3.02	3.14
007900	Dried fruit mix with nuts	0.14	0.14	3.00	2.97
011800	Potatoes	0.13	0.14	2.73	2.84
000700	Wholemeal bread	0.13	0.13	2.70	2.66
009200	Bananas	0.13	0.12	2.66	2.53
002000	Wheat flour	0.10	0.12	2.18	2.52
012800	Fruit drops, jelly	0.12	0.12	2.49	2.42
002400	Curd cheese cake	0.11	0.11	2.21	2.37
002100	Pasta	0.12	0.11	2.60	2.33
008500	Apples	0.11	0.11	2.38	2.23
007600	Pure vegetable oil	0.10	0.11	2.16	2.23
010900	Iceberg lettuce	0.10	0.10	2.12	2.10
011000	Paprika	0.09	0.10	1.90	2.09
001550	Cake	0.09	0.09	1.92	1.90
008900	Grapes	0.08	0.09	1.76	1.89
011700	Onions	0.08	0.08	1.62	1.76
010200	Potatoe chips	0.08	0.08	1.77	1.73
000100	Pizza, deep frozen	0.07	0.07	1.52	1.52
013800	Red pepper	0.07	0.07	1.57	1.50
013200	Crystal sugar	0.05	0.07	1.11	1.50
001400	Cereals	0.08	0.07	1.66	1.48
008800	Peaches, nectarines	0.07	0.07	1.55	1.44
008400	Strawberries	0.07	0.07	1.37	1.40
010600	Cucumbers	0.06	0.07	1.23	1.36
010800	Carrots	0.06	0.06	1.25	1.24
009400	Oranges	0.06	0.06	1.17	1.20
001110	Lye bun	0.05	0.06	1.13	1.18
011220	Packed salad	0.06	0.05	1.22	1.12
001100	Half baked buns	0.05	0.05	1.10	1.11
010700	Cauliflower	0.06	0.05	1.30	1.10
001500	Butter biscuits	0.06	0.05	1.18	1.09
001900	Long grain rice	0.05	0.05	1.10	1.04
009900	Mixed vegetables, deep frozen	0.05	0.05	0.97	1.03
008200	Tangerines	0.04	0.05	0.93	1.03
007800	Olive oil	0.06	0.05	1.18	1.02
009500	Lemons	0.05	0.05	1.14	1.00
010500	Champignons	0.04	0.04	0.89	0.89
009000	Musk melons, cantaloupes	0.05	0.04	1.05	0.89
010100	French fries, deep frozen	0.04	0.04	0.86	0.86
008440	Berries	0.05	0.04	0.97	0.85
001200	Ready-made dough	0.03	0.04	0.70	0.75
013000	Jam	0.03	0.04	0.72	0.73
001600	Wafers with hazelnut cream	0.04	0.03	0.75	0.71
008100	Raisins	0.04	0.03	0.75	0.66
012300	Tinned peaches	0.03	0.03	0.65	0.62
010000	Spinach, deep frozen	0.03	0.03	0.53	0.60
007400	Margarine	0.03	0.03	0.64	0.59
010490	Avocado	0.03	0.02	0.53	0.52
013100	Natural honey	0.02	0.02	0.49	0.48
008000	Peanuts salted	0.02	0.02	0.49	0.48
008600	Pears	0.03	0.02	0.53	0.47
009300	Kiwi	0.02	0.02	0.42	0.43
000200	Yeast dumpling, deep frozen	0.02	0.02	0.38	0.43
001700	Salty sticks	0.02	0.02	0.40	0.40

Table 6: Animal-based foods (weights in percent).

Code	Name	Basket weight 2022	Basket weight 2023	Target weight 2022	Target weight 2023
004200	Pork ham	0.23	0.23	5.27	5.12
005660	Extended shelf life milk	0.19	0.21	4.35	4.75
006500	Eggs	0.21	0.21	4.82	4.71
004000	Sausages	0.21	0.20	4.71	4.48
005000	Chicken breast	0.17	0.19	3.97	4.13
006700	Gouda	0.16	0.18	3.70	3.95
007300	Butter	0.16	0.17	3.67	3.74
002600	Minced meat	0.14	0.16	3.24	3.65
004400	Pork sausage	0.14	0.16	3.29	3.48
006600	Emmentaler	0.13	0.14	2.98	3.04
004800	Turkey breast	0.11	0.12	2.60	2.62
003700	Pork cutlet	0.12	0.12	2.75	2.56
003800	Pork sirloin	0.11	0.11	2.54	2.44
002800	Bacon	0.11	0.11	2.59	2.42
003100	Roast beef	0.11	0.10	2.50	2.33
004600	Salami	0.10	0.10	2.36	2.32
005700	Milk shake	0.11	0.10	2.54	2.31
004700	Roast chicken	0.09	0.10	2.13	2.29
003200	Beef round	0.10	0.10	2.29	2.21
006800	Hard cheese	0.10	0.10	2.26	2.18
005400	Fresh fish	0.10	0.10	2.21	2.14
004100	Dry sausage	0.10	0.10	2.21	2.12
006200	Yoghurt fruit flavoured	0.09	0.09	2.04	2.09
005900	Whipped cream	0.08	0.08	1.86	1.88
006400	Yoghurt	0.09	0.08	1.94	1.87
007200	Mozzarella	0.07	0.08	1.65	1.74
003300	Beef shoulder	0.08	0.08	1.83	1.73
006900	Camembert	0.08	0.08	1.80	1.73
003600	Pork chops	0.08	0.08	1.77	1.70
004300	Turkey sausage	0.06	0.08	1.43	1.69
005100	Tunafish	0.08	0.07	1.79	1.64
007100	Fresh cream cheese	0.06	0.07	1.35	1.51
005800	Sour cream	0.06	0.07	1.40	1.48
005200	Codfish filet, deep frozen	0.07	0.07	1.57	1.46
005300	Fish fingers, deep frozen	0.06	0.06	1.30	1.41
002700	Smoked meat	0.05	0.05	1.21	1.19
005661	Fresh milk	0.05	0.05	1.22	1.18
003000	Beef goulash canned	0.04	0.05	0.96	1.09
002900	Liver pasty	0.05	0.04	1.03	1.00
005500	Smoked salmon	0.05	0.04	1.09	0.96
004900	Breaded chicken meat, deep frozen	0.04	0.04	0.92	0.91
003500	Pork belly	0.04	0.04	0.84	0.82
006100	Evaporated milk	0.04	0.04	0.86	0.82
006000	Curd cheese	0.02	0.03	0.57	0.60
002500	Veal cutlet	0.03	0.02	0.61	0.53

Table 7: Miscellaneous food products and beverages (weights in percent).

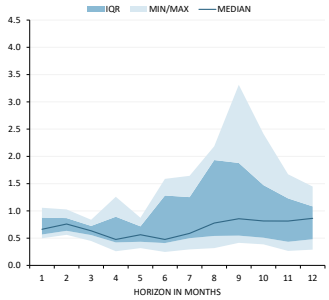
Code	Name	Basket weight 2022	Basket weight 2023	Target weight 2022	Target weight 2023
		Miscellaneous food products			
012550	Chocolate box	0.14	0.14	14.31	14.11
014101	Convenience food, chilled	0.13	0.13	13.38	13.65
012200	Ice cream, family size	0.13	0.13	13.39	13.04
012500	Milk chocolate	0.12	0.11	11.59	10.85
012700	Chewing gum	0.09	0.09	9.35	9.59
014103	Baby food (milk)	0.07	0.07	7.24	7.50
014000	Ketchup	0.06	0.06	5.74	6.04
000300	Convenience food, deep frozen	0.05	0.05	4.99	5.11
013900	Mustard	0.05	0.05	4.68	5.01
014100	Vinegar	0.05	0.05	5.05	5.00
013400	Soup powder	0.05	0.05	4.99	4.96
012600	Chocolate bar	0.03	0.03	3.38	3.21
013700	Salt	0.02	0.02	1.92	1.93
		Alcoholic beverages			
016300	Bottled beer	0.35	0.35	22.39	23.00
015800	White wine	0.31	0.30	20.04	19.49
016100	Canned beer	0.27	0.27	17.39	17.30
015700	Red wine	0.19	0.18	11.98	11.82
016000	Sparkling wine	0.13	0.13	8.21	8.15
015500	Vodka	0.10	0.11	6.60	6.96
015400	Rum	0.07	0.07	4.31	4.38
015550	Liqueur	0.07	0.07	4.34	4.32
016250	Beer, special	0.04	0.04	2.63	2.59
016400	Mixed beer	0.03	0.03	2.12	1.98
		Non-alcoholic beverages			
014300	Coffee	0.20	0.21	15.95	16.94
015000	Cola	0.16	0.15	12.92	12.60
014800	Mineral or table water	0.14	0.15	11.44	12.09
014301	Coffeepads, caps	0.15	0.13	12.31	10.99
015200	Orange juice	0.11	0.12	9.19	9.74
014900	Soft drink carbonated	0.10	0.10	7.94	8.04
015300	Apple juice	0.09	0.08	7.04	6.95
014700	Energy drink	0.08	0.08	6.79	6.64
014200	Tea in bags	0.08	0.07	6.44	5.81
014302	Instant coffee	0.06	0.06	5.14	5.19
014600	Mineral water, flavoured	0.04	0.04	3.06	3.36
014500	Cocoa instant drink	0.02	0.02	1.77	1.65

B Model selection

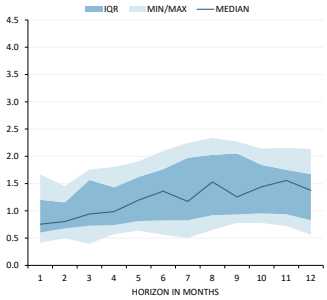
Figure B.1: Best models by forecast year and horizon

	Horizon	PLANT	ANIMAL	MISC	ALCOHOLIC	NON-ALCOHOLIC
2018	1	BART – Set 1	BART – Set 2	Lasso – Set 2	Elastic Net – Set 3	Random Forest – Set 3
	2	ELM Lasso – Set 1	BART – Set 2	Hansen Select – Set 1	Lasso – Set 3	BART – Set 2
	3	ELM Linear – Set 2	Lasso – Set 3	BART – Set 2	Hansen Combine – Set 1	BART – Set 2
	4	ELM Linear – Set 2	Random Forest – Set 1	Ridge – Set 2	Lasso – Set 2	BART – Set 2
	5	ELM Step – Set 3	Random Forest – Set 2	Random Forest – Set 3	ARMA	Random Forest – Set 2
	6	ELM Step – Set 3	Random Forest – Set 2	Ridge – Set 3	Hansen Combine – Set 1	Elastic Net – Set 3
	7	MPL – Set 1	Lasso – Set 1	Lasso – Set 2	ARMA	Random Forest – Set 3
	8	ELM Linear – Set 1	Lasso – Set 1	Random Forest – Set 1	ARMA	Random Forest – Set 3
	9	ELM Step – Set 3	Ridge – Set 1	ARMA	ARMA	Elastic Net – Set 3
	10	ELM Step – Set 2	BART – Set 2	Random Forest – Set 1	ARMA	Elastic Net – Set 2
	11	ELM Step – Set 1	Random Forest – Set 3	Random Forest – Set 1	ARMA	Lasso – Set 3
	12	ELM Linear – Set 1	Elastic Net – Set 2	Random Forest – Set 1	ARMA	ELM Lasso – Set 1
2019	1	ARMA	ELM Lasso – Set 2	Hansen Select – Set 1	ELM Linear – Set 2	ARMA
	2	ELM Linear – Set 3	ELM Ridge – Set 3	Ridge – Set 1	ARMA	ARMA
	3	ELM Step – Set 2	ELM Ridge – Set 2	Random Forest – Set 2	ELM Step – Set 3	ARMA
	4	ELM Linear – Set 2	ELM Lasso – Set 3	BART – Set 3	ELM Step – Set 3	ARMA
	5	ARMA	Random Forest – Set 1	Elastic Net – Set 1	ELM Step – Set 1	ARMA
	6	ARMA	BART – Set 2	ELM Ridge – Set 1	ELM Lasso – Set 3	BART – Set 3
	7	ARMA	Random Forest – Set 1	Lasso – Set 3	Lasso – Set 3	ELM Lasso – Set 2
	8	ELM Step – Set 3	Random Forest – Set 1	Ridge – Set 1	ELM Ridge – Set 2	ELM Ridge – Set 3
	9	MPL – Set 1	Random Forest – Set 1	Lasso – Set 1	ELM Ridge – Set 2	Elastic Net – Set 3
	10	MPL – Set 1	Random Forest – Set 3	Ridge – Set 1	ELM Ridge – Set 3	Random Forest – Set 3
	11	ELM Lasso – Set 1	Random Forest – Set 1	ARMA	ELM Ridge – Set 2	Lasso – Set 3
	12	ELM Linear – Set 1	Lasso – Set 2	ELM Linear – Set 2	ELM Lasso – Set 3	Elastic Net – Set 2
2020	1	BART – Set 2	Random Forest – Set 1	ELM Step – Set 2	BART – Set 3	ELM Ridge – Set 2
	2	BART – Set 2	Random Forest – Set 1	Hansen Combine – Set 1	Random Forest – Set 1	ELM Ridge – Set 2
	3	ARMA	Ridge – Set 1	ELM Linear – Set 2	Lasso – Set 3	Ridge – Set 1
	4	ARMA	Random Forest – Set 2	Random Forest – Set 2	Hansen Select – Set 1	Random Forest – Set 2
	5	Ridge – Set 2	Elastic Net – Set 1	ELM Ridge – Set 1	Hansen Select – Set 1	BART – Set 2
	6	BART – Set 1	Random Forest – Set 3	ELM Ridge – Set 1	Hansen Select – Set 1	Ridge – Set 1
	7	MPL – Set 1	Random Forest – Set 3	ARMA	Hansen Select – Set 1	Lasso – Set 1
	8	Elastic Net – Set 2	Random Forest – Set 3	ARMA	Hansen Select – Set 1	Random Forest – Set 3
	9	Random Forest – Set 2	Random Forest – Set 2	ELM Ridge – Set 1	Hansen Combine – Set 1	Random Forest – Set 2
	10	Random Forest – Set 2	BART – Set 1	MPL – Set 1	Hansen Combine – Set 1	ELM Linear – Set 1
	11	Random Forest – Set 2	BART – Set 1	ELM Linear – Set 1	Hansen Select – Set 1	ELM Lasso – Set 2
	12	Random Forest – Set 2	ELM Ridge – Set 1	MPL – Set 1	MPL – Set 1	MPL – Set 1
2021	1	ELM Linear – Set 3	BART – Set 3	ELM Linear – Set 2	Elastic Net – Set 2	ELM Linear – Set 3
	2	ELM Ridge – Set 1	Lasso – Set 1	Elastic Net – Set 2	Elastic Net – Set 2	Lasso – Set 3
	3	Random Forest – Set 1	Hansen Combine – Set 1	ELM Step – Set 3	Random Forest – Set 3	ELM Ridge – Set 1
	4	ELM Lasso – Set 1	Elastic Net – Set 1	Hansen Combine – Set 1	BART – Set 3	Random Forest – Set 1
	5	Random Forest – Set 3	Random Forest – Set 1	Elastic Net – Set 2	Lasso – Set 2	Random Forest – Set 1
	6	Elastic Net – Set 3	BART – Set 1	Elastic Net – Set 1	Lasso – Set 2	BART – Set 1
	7	Random Forest – Set 3	Hansen Combine – Set 1	Ridge – Set 1	Ridge – Set 3	BART – Set 1
	8	BART – Set 2	BART – Set 1	Lasso – Set 1	BART – Set 2	BART – Set 1
	9	BART – Set 2	Ridge – Set 3	BART – Set 1	BART – Set 2	Random Forest – Set 3
	10	Hansen Select – Set 1	Ridge – Set 3	Elastic Net – Set 1	Hansen Select – Set 1	Lasso – Set 2
	11	Lasso – Set 1	Lasso – Set 3	Hansen Select – Set 1	MPL – Set 1	Random Forest – Set 2
	12	Random Forest – Set 1	Ridge – Set 3	Lasso – Set 1	MPL – Set 1	Random Forest – Set 2

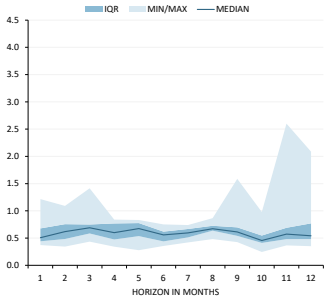
Figure B.2: The spread of mean absolute forecast errors



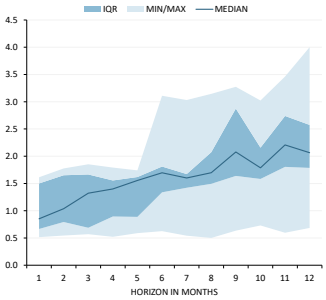
(a) Plant 2018



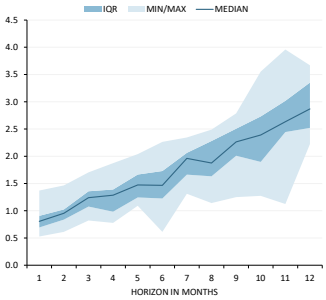
(b) Animal 2018



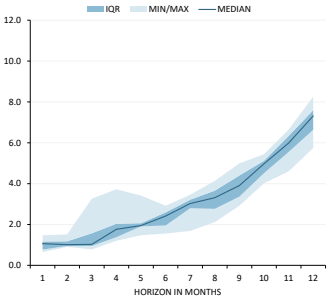
(c) Misc. 2018



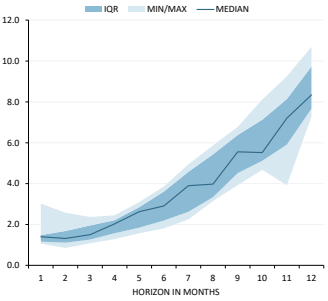
(d) Alcoholic 2018



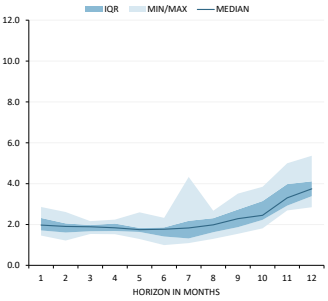
(e) Non-Alcoholic 2018



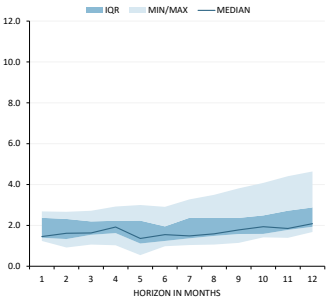
(f) Plant 2021



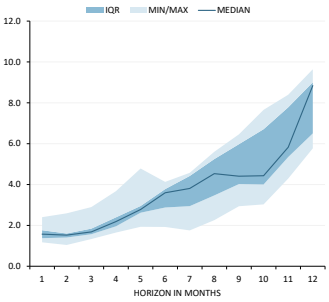
(g) Animal 2021



(h) Misc. 2021

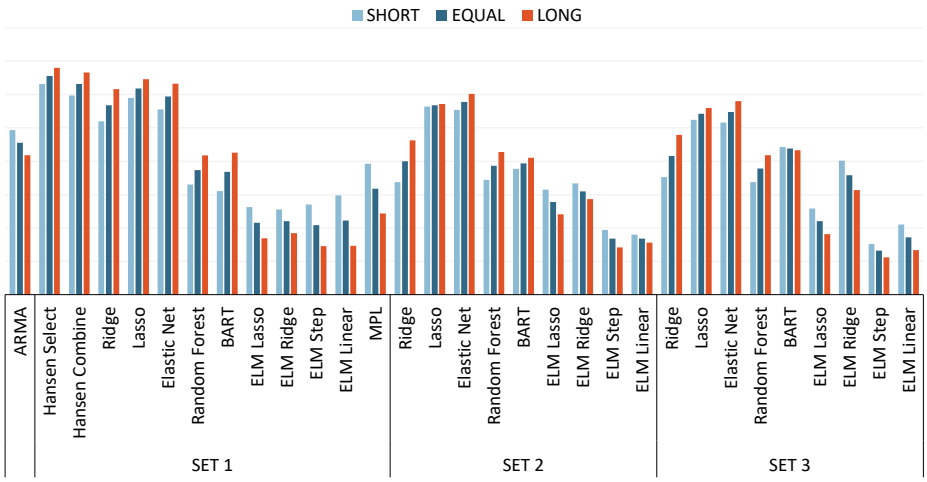


(i) Alcoholic 2021

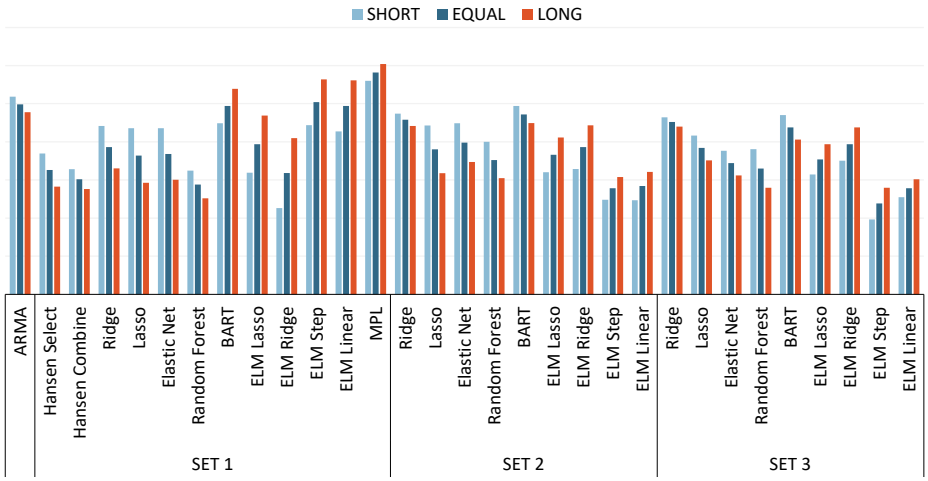


(j) Non-Alcoholic 2021

Figure B.3: Horizon-weighted model rankings (Plant)

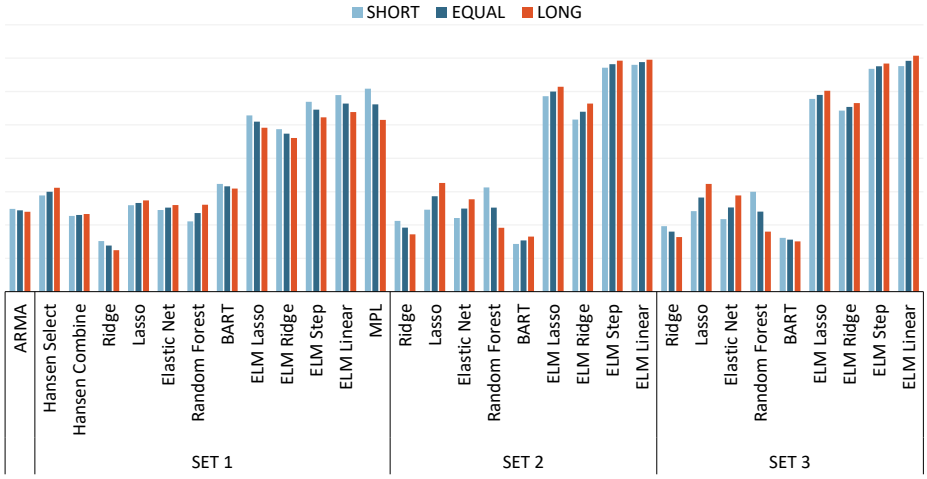


(a) 2018

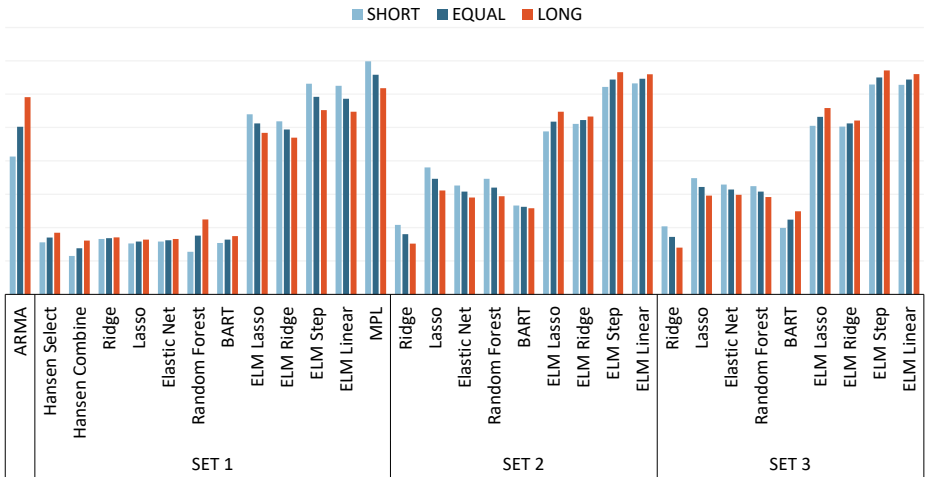


(b) 2021

Figure B.4: Horizon-weighted model rankings (Animal)

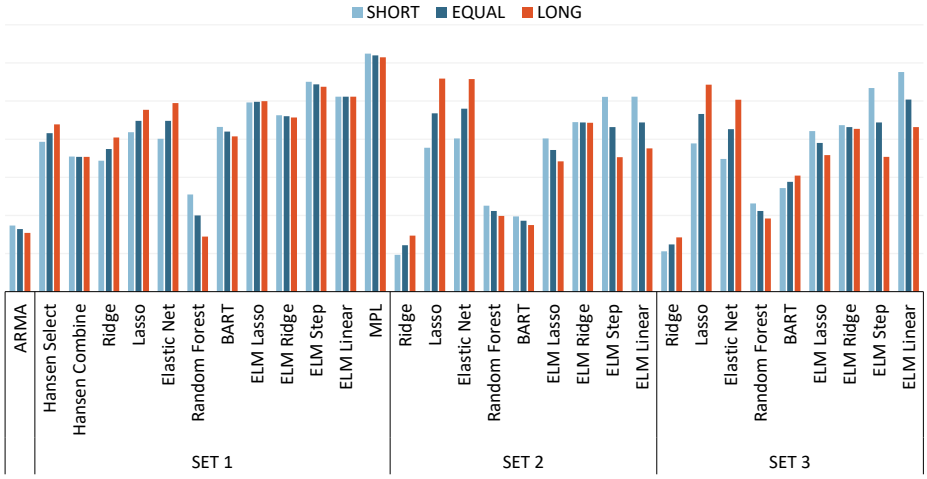


(a) Forecast year 2018

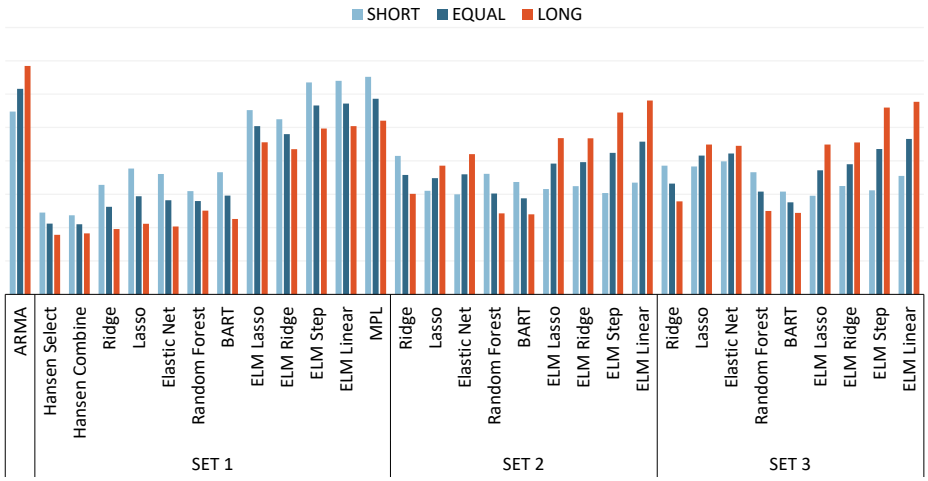


(b) Forecast year 2021

Figure B.5: Horizon-weighted model rankings (Misc.)

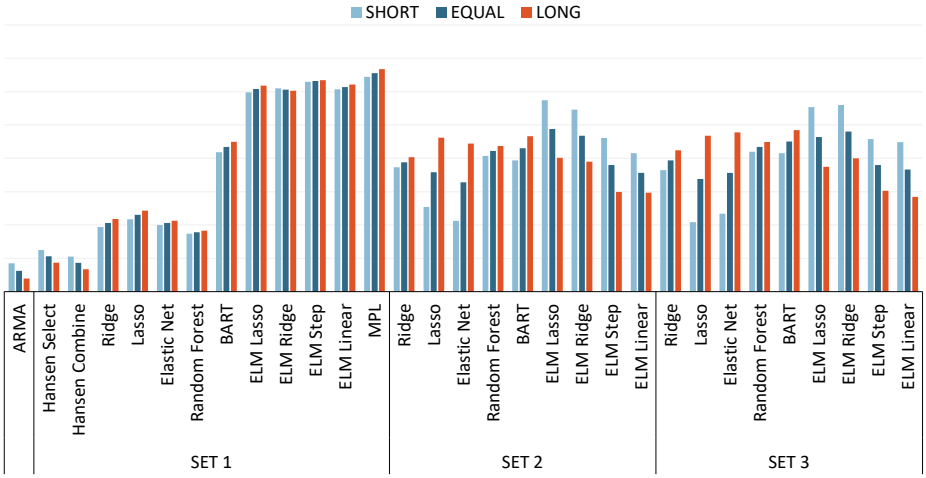


(a) Forecast year 2018

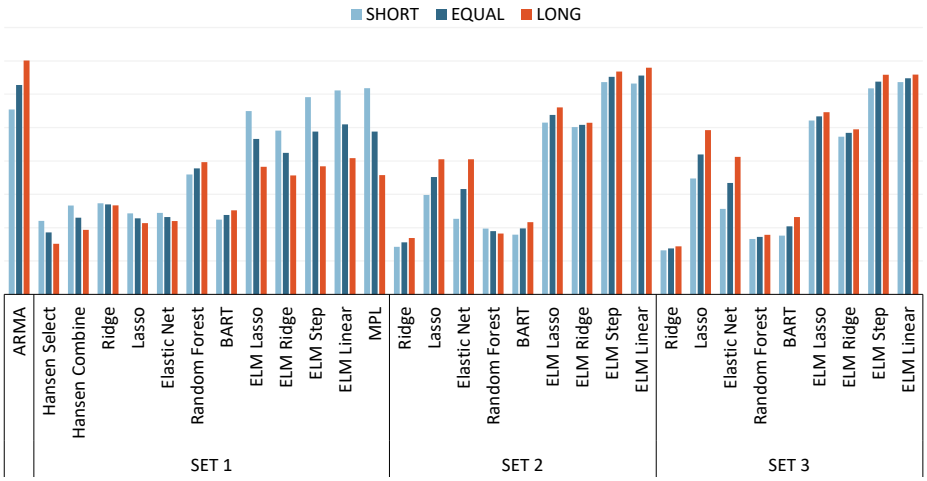


(b) Forecast year 2021

Figure B.6: Horizon-weighted model rankings (Alcoholic)

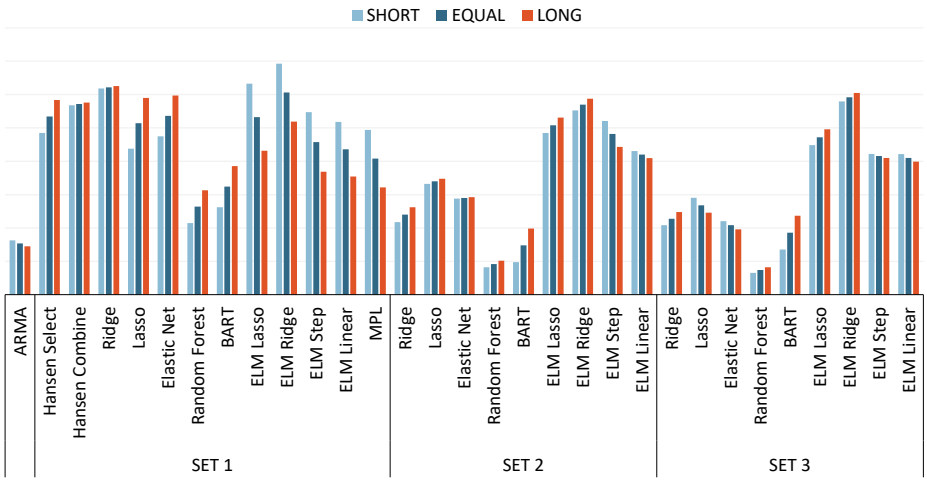


(a) Forecast year 2018

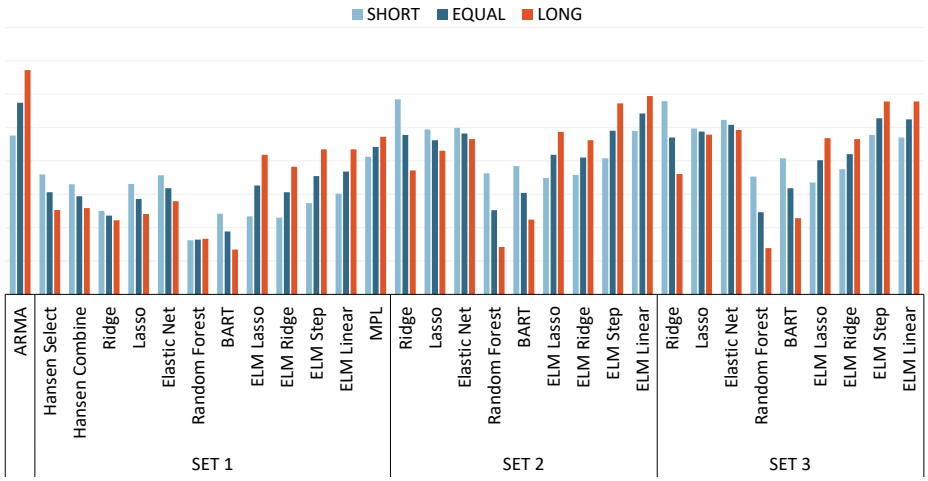


(b) Forecast year 2021

Figure B.7: Horizon-weighted model rankings (Non-alcoholic)



(a) Forecast year 2018



(b) Forecast year 2021