NBP Narodowy Bank Polski

NBP Working Paper No. 302

Food inflation nowcasting with web scraped data

Paweł Macias, Damian Stelmasiak

NBP Working Paper No. 302

Food inflation nowcasting with web scraped data

Paweł Macias, Damian Stelmasiak

Narodowy Bank Polski
Warsaw 2019

Paweł Macias – Narodowy Bank Polski; pawel.macias@nbp.pl
Damian Stelmasiak – Narodowy Bank Polski; damian.stelmasiak@nbp.pl

Acknowledgements
We thank Karol Szafranek and Grzegorz Szafrański for valuable comments. Helpful remarks from participants of the NBP Workshop on Forecasting and numerous research seminars at Narodowy Bank Polski are gratefully acknowledged. All errors are our own. The views and opinions presented in this paper are those of the authors and do not reflect the view of Narodowy Bank Polski.

Published by:
Narodowy Bank Polski
Education & Publishing Department
ul. Świętokrzyska 11/21
00-919 Warszawa, Poland
www.nbp.pl

ISSN 2084-624X

Contents

Abstract

In this paper we evaluate the ability of web scraped data to improve nowcasts of Polish food inflation. The nowcasting performance of online price indices is compared with aggregated and disaggregated benchmark models in a pseudo real-time experiment. We also explore product selection and classification problems, their importance in constructing web price indices and other limitations of online datasets. Therefore, we experiment not only with raw indices, but also with several approaches to include them into model-based forecasts. Our findings indicate that the optimal way to incorporate web scraped data into regular forecasting is to include them in simple distributed-lag models at the lowest aggregation level, combine the forecasts and aggregate them using statistical office methodology. We find this approach superior to other benchmark models which do not take online information into account.

JEL classification: E37, C81, C55

Keywords:

web scraping, nowcasting, inflation, big data, online prices

1 Introduction

Web scraping activities offer a relatively cheap solution to the demand for fast arriving price information for inflation measurement and forecasting. Unlike with survey data, we do not experience any delays in data collection, which allows us to prepare forecasts and analyses almost in real time. Moreover, we are not limited by third party services, which may limit access to micro data regardless of the costs or provide the data with delay, as is often the case with scanner data. The web scraped data have many desirable features – they often include not only prices but also information on discounts, product descriptions, and sometimes product availability across shop branches. With web scraping it is possible to fetch the data at any frequency – weekly or daily. Web scraped prices are perceived as useful in forecasting inflation (Powell et al., 2018) since they enable to utilize the most current retail price data, mimic official CPI price dynamics, monitor prices in real-time and analyse price rigidities at the product level.

The principal aim of this research is to verify whether web prices are helpful in nowcasting food inflation in Poland. Food inflation forecasts remain an important task in regular inflation forecasting as food constitutes a major part of consumer inflation – it currently accounts for 24% household expenditures in Poland. Taking into account high volatility of food prices and seasonal outliers, food inflation contributes significantly into headline inflation (see Fig. 1 in Szafranek and Hałka, 2017). However, there are evident obstacles to obtain high-grade data for food web prices as grocery items are not massively bought online. Not all major grocery retailers offer functional internet stores. This raises the question whether at current level of market development this source of data is ample enough to improve forecasts upon benchmarks. Nevertheless, it is expected that the quality of online price data will increase gradually due to the progress in digitalization of grocery retailing.

Growth of the e-commerce market in Poland has been very dynamic so far. The share of e-commerce in the turnover enterprises more than doubled in a decade and in 2017 it was estimated at 15% (Eurostat). The development of the e-grocery market has been even faster. The percentage of individuals who bought food online features a greater than 10-fold increase since 2005, and in 2017 it amounted to 11% (Eurostat). Complimentary surveys suggest that 28% of internet users in Poland have bought food online at least once and 16% buy on a regular basis (E-grocery in Poland report, 2017).

The impetuous expansion of the e-commerce grocery market is primarily limited by logistical constraints. For example, on the developed UK e-grocery market there are high fulfillment costs – an average cost of packing and delivering is higher than the delivery fee set for customers at ca. $7-$17 per order (Fung Global Retail & Technology report, 2016). Despite additional costs related to online sales, Euromonitor International anticipates an increase in e-grocery market by 11.2% in Eastern Europe and by 9.2% in Western Europe in 2015-2020, much higher figures than estimates of growth in store-based grocery.

Solid outlook for further expansion of e-commerce in the grocery market is well-founded also due to favorable purchase behavior patterns. It is reported that if a customer experiences an online purchase, the new behavior is usually retained in contrast to the non-grocery segment where shopping patterns revert to pre-online rather quickly (Melis, 2016).

The grocery market in Poland has exceptionally high share of superettes, traditional shops and small supermarkets, coupled with a low share of hypermarkets as compared to other European countries (Nielsen Grocery Universe 2017). Due to the low concentration of domestic retail market, in our web scraping process we cover relatively low market share. However, we expect that price growth rates are similar in traditional and online groceries in Poland, which enables us to forecast inflation

based on web scraped data. The sample used in this paper covers all major web stores in the Polish online grocery market.

The web scraped prices are selected, classified and aggregated to produce the total food price index. We assess the forecasting accuracy of online prices aggregate alone and within simple linear distributed-lag models and their combinations. We use the (pseudo) real-time scheme to prepare nowcasts, the monthly online price index is calculated in the middle of the month, just after the CPI is published. Forecast accuracy is measured by the root mean squared error in the period from January 2014 till June 2018 and it is compared to benchmarks that do not include web scraped data. Preliminary results suggest that employing web scraped data improves nowcasts with respect to the ARMA baseline model. The advantage over the benchmark increases considerably when the online price index is introduced into ADL models.

The paper is organized as follows: Section 1.1 is literature summary focused on forecasting applications of web scraped data. Section 2.1 introduces web-scraping methods used in our research. Sections 2.2 and 2.3 present, respectively, our database structure and dataset itself. Section 2.4 explains the advantages of proper product selection and classification. Section 2.5 presents the methodology of the out-of-sample inflation forecasting exercise for Poland carried out with the use of real-time online data. Section 3 reports the forecasting accuracy of pure online data indices and model-based approaches. Finally, section 4 summarizes our conclusions of online data usefulness and the optimal way to use them in regular forecasting.

2 Literature overview

The applications of online or web scraped data are in general three-folded and include inflation measurement, inflation forecasting (including nowcasting) and micro price setting mechanism researches.

It should be noted that only few researches on web prices have been carried out so far. An early contribution is due to Lunnemann and Wintr (2006) where they find differences in price stickiness between web and physical store prices in Europe and the USA. Then, in 2008, the Billion Prices Project was created at MIT. It has remained the largest project focused on web scraping and online prices analysis till now (Cavallo and Rigobon, 2016). Huge amounts of data downloaded every day primarily make it possible to calculate CPI-like price indices. For example, Cavallo (2013) finds that the inflation measure based on web prices is similar to official headline inflation in Latin America except for Argentina. Secondly, the data make other price setting policy studies possible, e.g. price stickiness evaluation, online vs offline price synchronization issues (Cavallo, 2017), impacts of government price controls (Aparicio and Cavallo, 2018), etc. Finally, daily web scraped prices may be compared to other data sources like scanner data and official CPI data to assess measurement bias and to better understand price setting mechanism (Cavallo, 2018).

In their seminal paper, Bertolotto et al. (2014) show that web scraped data are useful in forecasting CPI including nowcasting and longer horizons. This research is to our knowledge the first successful attempt to forecast the broad CPI (full basket of products) using web scraped data. In Europe, a rather early web scraping pilot was started by ONS (Swier, 2014, see details in Breton et al., 2016, Bhardwaj et al. 2017) in January 2014 as a part of ONS Big Data Project. It is reported to initially support 3 supermarkets operating in the UK and uses scraping routines written in Python. One of the reasons for starting web scraping activity was the low availability of scanner data (Breton et al., 2016), which still remains a problem for many countries.

However, they do not share their experiences with forecasting using this data. Radzikowski and Smietanka (2016) publish an online-based CPI-like price index for Poland, although they do not provide any details on forecasting.

Central banks, which regularly produce nowcasts and forecasts of inflation are also increasingly interested in utilizing online prices. The Central Bank of Armenia collects online food prices in order to produce flash estimates and forecasts of food inflation (Aghajanyan et al., 2017). Researchers from Riksbank, Hull et al. (2017), present forecasting results for selected items of food prices (fruit and vegetables) in Sweden. They indicate online prices aggregates beat (in terms of RMSE) official Riksbank nowcasts of the fruit and vegetables index.

Aparicio and Bertolotto (2017) continue the work of Bertolotto et al. (2014). Forecasts from their model enriched with online prices beat simple benchmarks and two leading surveys of professional forecasters. Despite the fact that the high frequency information advantage is not fully used (as they are using simple linear models with no mixed frequency data plugged) the approach still provides significant improvement over not using online data – even in the case where the latest online data portion is rejected. The hypothesis is that online prices are adjusted more frequently than offline prices, so official statistics possibly experiences some delays in capturing real world price dynamics.

Powell et al. (2018) use web scraped data in forecasting daily log-prices of selected food and alcoholic beverages. They find web scraped data useful in daily forecasting exercise – they report up to 30% reduction in RMSFE over the benchmark for some categories (almost 10% reduction on average for selected food and non-alcoholic beverages groups).

Irrespective of using online price datasets, there are many approaches that are considered beneficial particularly in inflation forecasting. Forecasting disaggregated price indices can significantly improve forecasts if an appropriate model is applied

(Bermingham and D'Agostino, 2011; Huwiler and Kaufmann, 2013). However, it is not an easy task to find proper model specification for each inflation component. Therefore to reduce model selection bias the combination of forecasts is often considered.

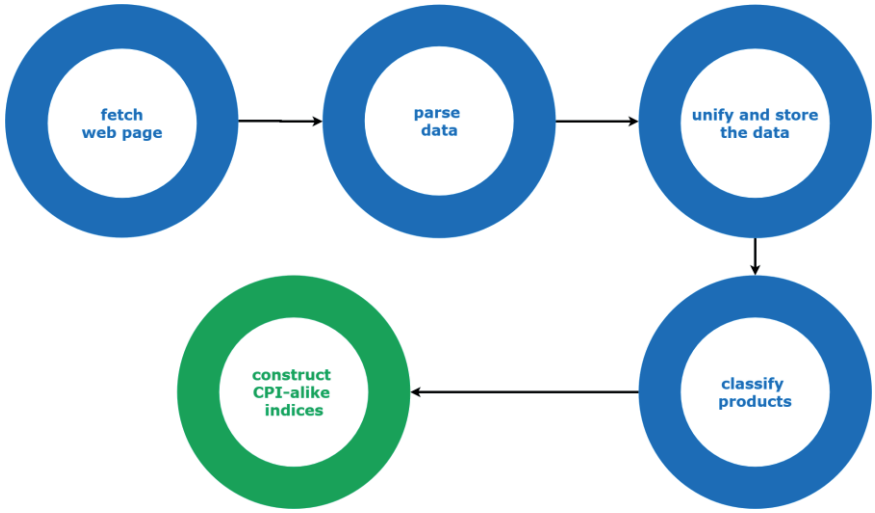
Faust and Wright (2013) report the results of comprehensive comparison of models including Phillips curve, DSGE, factor model, Bayesian approaches and it should be noted that simple benchmarks like AR(1) are still hard to beat. Szafranek (2017) also finds that more sophisticated models not always outperform random walk benchmarks in forecasting of the Polish CPI.

3 Methodology

In November 2009 the eCPI Project was started in Narodowy Bank Polski, aimed at collecting prices from internet shops in Poland. Starting the project we aimed at constructing food and non-alcoholic beverages index based on fast-arriving online prices. Contrary to other available data that are useful in forecasting (like agricultural commodity prices), the eCPI is believed to be a direct source of information on retail prices. While the project is still focused on groceries, in 2017 we started scraping clothing, footwear, home-improvement stores as well as airplane tickets. We continue expanding the list of stores being scraped to cover possibly the largest part of household expenditures.

Prices in web stores are published in a very distinctive way as stores' main objective is to create user friendly platform for buying their products. Therefore, prices are among of other plenty elements in web page, distributed among various category trees, pages and sub-pages. Most often we need to deal with unstructured or very loosely structured data, which pose a technical challenge. Since there is no easy way to download data from online stores it is necessary to use programming techniques to retrieve them from web pages. Most often the process of obtaining data is as follows (see Fig. 1). In the first step we fetch all web pages of a given online store, which contain information about products. However, data of our interest are still embedded in the web page source code, so they are poorly structured and require further processing. Therefore, in the second stage, the downloaded data are parsed. This simply means that we identify and extract the data on prices and product features from the full, patchy web page codes. We also check the correctness of the data acquisition process. At the last stage, the data are unified and the scraped products are selected and classified. A newly created, unified database is ready for forecasting and other applications.

Fig. 1 Schematic process of obtaining data.



Web scraping is conducted in Python using Selenium, Requests, Beautiful Soup as well as auxiliary libraries to fetch and process data. We collect online prices every day using a cloud server. We perform data acquisition while minimizing the burden on web stores owner and trying to be in line with the Code of Practice for Statistics (see Greenaway, 2018) by *delaying accessing pages on the same domain and scraping at a time of day when the website is unlikely to be experiencing heavy traffic.* Additionally, in most cases we choose the most effective way possible to scrape the data (see the scraping techniques discussion below and Tab. 1) and minimize server traffic.

3.1 Web scraping techniques

We distinguish three main web scraping methods, which we use on a daily basis: parsing raw web page sources, interacting with Document Object Model (DOM) in live web browsers or direct fetching of structured data with Application Programming Interface (API). The web page source approach is based on obtaining the web page source and parsing (extracting) information from the HTML or JavaScript tags. To obtain the page source, one may use any method of downloading it like the headless session by means of Requests library or just save it after the site is loaded in an emulated web browser. In this approach HTML tag parsers are

extensively used and sometimes regular expressions that search for hard-coded JSON structures are helpful. The DOM-object approach relies on interacting with rendered objects on the page, which is possible in web browser only – we use test automation software for web browsers. Dynamic JavaScript web pages are handled properly because the web page is fully executed in exactly the same manner when every user accesses it. The API/direct connection to a store database consists in accessing publicly available well-structured JSON files that contain product details – they are obtained using public API or AJAX queries. However, most of APIs are strictly private or limited by other factors, therefore this approach is rarely seen in web scraping practice. We can also consider mixed strategies that combine elements of the aforementioned three approaches. Starting a web browser session and loading cookies into a headless session is a very useful method if non-default page settings are needed.

Tab. 1 Comparison of web scraping techniques.

Method	Speed	Processing	Stability	Availability	Data amount
Page source	Fast	Difficult	Good	High	sometimes more than on the screen
DOM-object	Slow	Difficult	Medium	Very high	on the screen
Direct/API	Very fast	Barely none	Very good	Low	often more than on the screen

Note: based on Authors' experience.

We find the direct method (API) to be of the highest quality overall although rarely available (Tab. 1). If there is no direct method available, we prefer the page source technique due to better speed and executing stability. In some cases of dynamic Javascript-based web pages the only way to get a product price is by the use of web browser automation extension and live interaction with DOM-objects. Nevertheless, this technique is the slowest and error-prone as we need to account for the page scripts loading time which varies.[1] Looking for speed and stability we often resort to

[1] Since June 2017 Google Chrome officially supports the headless mode (starting with version 59) and Firefox does so for Windows since September 2017 (version 56). The headless mode may increase speed and stability of web scraping when using web browsers.
https://developer.mozilla.org/en-US/Firefox/Headless_mode

mixed strategies – the most typical case is to properly prepare a web store session in a live web browser (i.e. setting the number of products per page) then to export the settings into a lightweight tool for web page source fetching.

Web scraping discussion

A bad choice of environment (and programming language) may lead to some inefficiencies, however, we find the choice of a suitable scraping technique (see Tab. 1) more important. According to Breton et al. (2016) "*Python, is not well suited to scraping websites that contain much JavaScript content*". In our opinion the problem like this does not occur because of the choice of the programming language but because of the use of inadequate web scraping methods. When a shop moves to a more dynamic JavaScript-based layout, then it needs a DOM-object based method, like web browser automation techniques that are general enough to handle those issues. The 'infinite scrolling problem' raised therein, to our knowledge, cannot occur in a web browser automation scenario as it would pose a problem to real visitors (customers). Therefore, we find Breton et al. (2016) criticism of particular language usage in web scraping largely exaggerated, while they do not discuss scrape approaches in deep.

In practice most of the efforts are focused on web scraping monitoring being done by a human on a daily basis. Our web scraping monitoring routine consists of checking the error logs and the size of parsed result files, which we find to be a good proxy to evaluate the corectness of scraping execution. In order to perform it faster and more easily, we use several tools dedicated to visualize results. The information about critical errors caused by web store pages modifications are handled through messenger app notifications for quicker reaction and code repair (update). Data as well as the dashboard visualization of data collection and error logs are saved in a cloud storage, which enables access on mobile devices. Our experience suggests that

https://developers.google.com/web/updates/2017/04/headless-chrome

faith in the 'extreme automation' feature of the web scraping process as described by Buono et al. (Eurostat, 2017) is illusory. Especially as for most web pages (when page source parsing and DOM-object based methods are considered) the web scraping process requires necessary code updates and adjustments since the web page structure might change any time. Techniques used in web scraping imply that the data collection is relatively cheap and easy but needs continuous monitoring and does not lead to full automation.

3.2 Database structure

The eCPI project is composed of several abstract layers that serve collecting, analytical and forecasting purposes. As our key objective in this paper is to nowcast Polish food inflation, we present the eCPI system only briefly.

The eCPI system is built upon a semi-distributed database, henceforth eCPIDB. The eCPIDB is stored in two forms: 1) a data lake (loose no-SQL structure) of daily web scraped data including both raw and initially processed (parsed) information from web stores, and 2) a time-frequency reduced relational database of monthly price averages as a convenient tool for macroeconomic analysis.

The advantage of the data lake is the much higher capacity compared to SQL-like databases, as an increasing number of web stores implies the necessity of concurrent web page fetching. At the time of writing this article the eCPIDB handles 6GB of data flowing every day from 22 stores (or ~200MB of pre-processed and ultra-compressed data per day). The data lake can also be easily put into HDFS for needs of distributed computations in Hadoop. We leave aside intermediate solutions like the MongoDB as we find the no-SQL database architecture ill-suited to high level analysis of results while the data inflow is not a constraint from the ex-post analysis perspective.

On the other hand, the reduced relational database of products and prices is relatively small and contains less than 20 GB of data. The database is updated a few

times a month and the data are unified and well structured. It allows a relatively easy and fast access, selection and classification as well as forecasting exercises or other macroeconomic applications.

3.3 Data

The data on web scraped grocery prices used in this paper span the period from Dec 2009 till Jun 2018. Until Dec 2016 we collected data on a weekly basis and currently we collect online prices every day. The eCPIDB contains information of over 75 million observations of food prices that cover 488,918 products in 4-7 grocery shops in Poland (see Fig. 2 and Tab. 2). There is a moderate variation in range of products measured by number of unit products[2] across the stores (see Fig. 3).

Fig. 2 Number of products and stores in eCPI database.

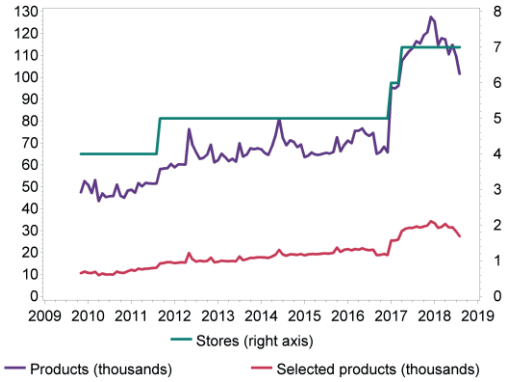
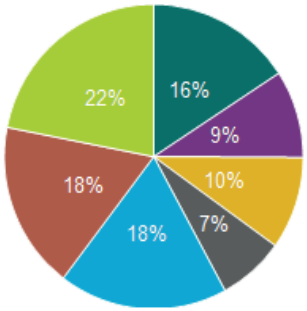


Fig. 3 Composition of products by store.



Note: The colours correspond to undisclosed online retailers, which we indicate by 1, 2, ..., 7.

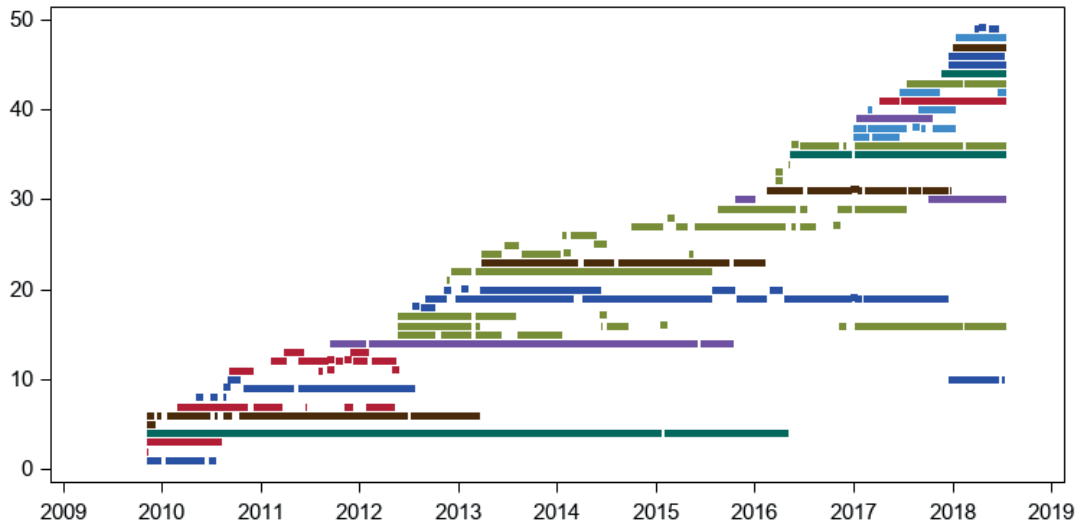
[2] By a product we denote an item which is exactly identified. Note that products are easily identifiable by their unique id or product description given by every web store. In general, the exact matching of products scraped from different stores is not possible. Therefore, some products from different shops are treated as separate products while they are indeed identical.

Tab. 2 Number of products and prices in eCPI database.

Store	Products	Selected products	Prices	Average No. of prices per day	No. of days price is observed - mean	No. of days price is observed - median
1	77 168	17 300	8 307 659	9 099	587	301
2	45 502	12 356	6 690 486	8 219	416	262
3	48 020	7 476	10 686 946	19 609	220	202
4	35 059	11 482	6 084 202	10 472	475	319
5	89 314	15 034	20 361 216	22 204	784	567
6	86 754	25 710	10 328 660	11 388	688	459
7	107 101	23 391	12 670 401	16 182	366	196
Total	488 918	112 749	75 129 570	97 173	534	322

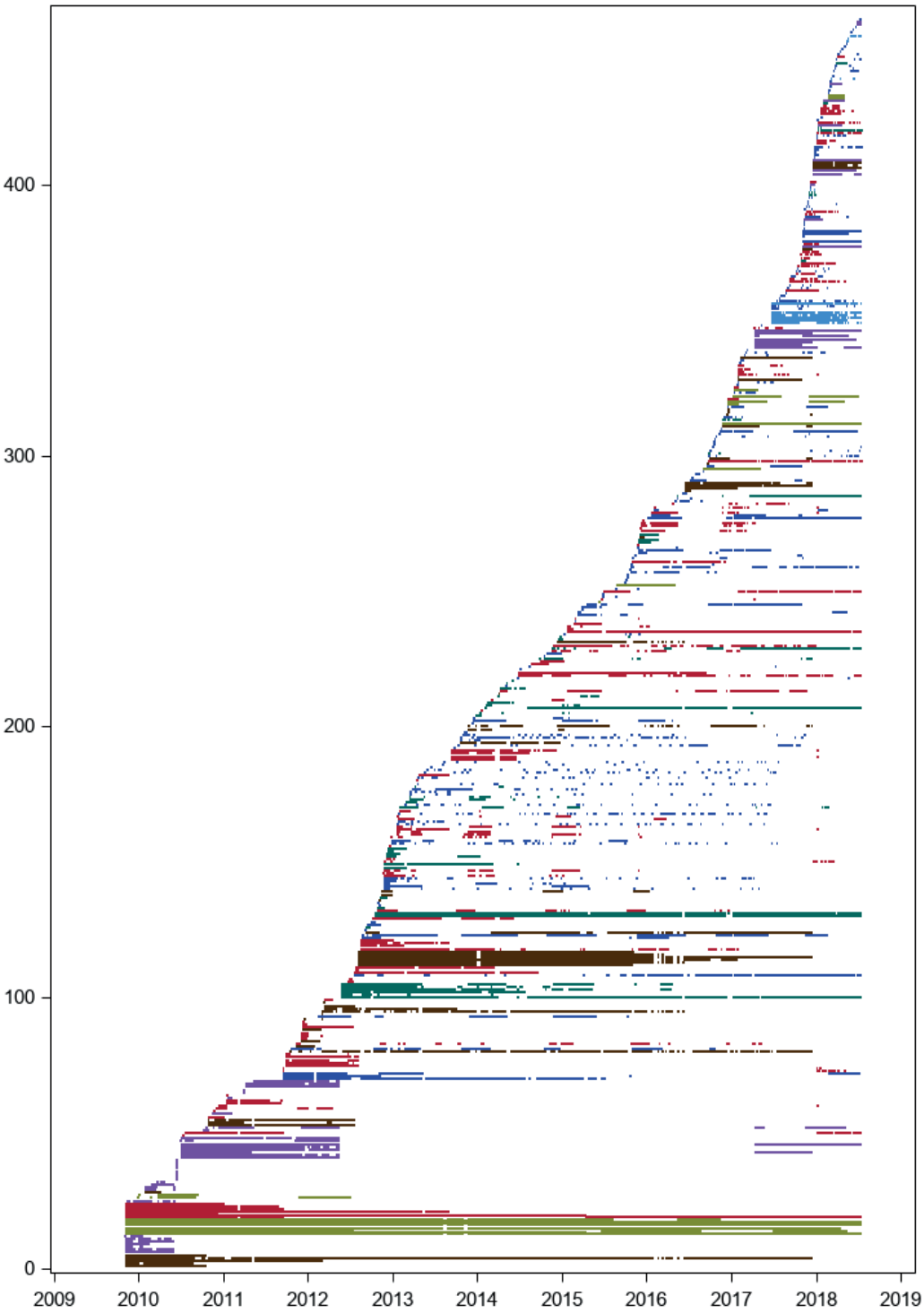
After meticulous selection, we usually obtain from 10 to 30 thousands unique products each day, however, the number of goods and stores varies over time, which is caused by several reasons. Firstly, the e-commerce market constantly grew during our research, so new stores kept appearing or expanding their offer. When a new significant store appeared on the market, it was being added to the eCPI project. Secondly, manufacturers often change the size or composition of their products, which implies the appearance of new products and the disappearance of previous ones. There are frequent promotions (e.g. products with an additional free product or in a bigger size container) or short series of products issued to check the preferences of customers. Thirdly, stores sometimes change the names of products (e.g. changing the word order, unit of size or adding additional information), thus it becomes more difficult to identify and track the same product in time. Fourthly, there might be errors in web scraping caused by change of store website or connection problems (e.g. a disabled website due to the maintenance of the server). Moreover, some products might be unavailable because of stock depletion. For the above reasons prices were observed for only 534 days on average (with the median being 322 days), see Tab. 2. All of these features are typical for web scraped data and result in product churn – see the examples of sugar and citrus fruits (Fig. 4 and Fig. 5).

Fig. 4 Product churn in sugar coloured by store.



Note: The figure shows a lifespan of products, stacked one by one. Each horizontal line represents the lifespan of one product, which may appear any time or when the web scraping of a specific store starts. Empty spaces correspond to product unavailability or technical problems with web scraping. Colours indicate different online stores.

Fig. 5 Product churn in citrus fruits coloured by store.



Note: The figure shows a lifespan of products, stacked one by one. Each horizontal line represents the lifespan of one product, which may appear any time or when the web scraping of a specific store starts. Empty spaces correspond to product unavailability or technical problems with web scraping. Colours indicate various online stores.

For the purpose of tracking inflation and forecasting food inflation we aggregate the data into monthly time series following the statistical office methodology. In the first step, we calculate the average price in the given month for each product as well as monthly growth rates. Disregarding missing daily data, we do not use any imputation methods. Neither do we use any kind of filters[3] to detect and exclude outliers as we do not find them beneficial in terms of error reduction in our dataset. In the next step, we calculate the average monthly growth rate by product groups using the geometric mean. Based on m-o-m dynamics we calculate other indices and aggregates.

In the literature one can find more refined methodologies of price index calculations, mostly employed by national statistical offices. However, their main objective differs from forecasting purposes as they see web scraped data as a possible way to enhance CPI or other price indices with high volume and affordable data, for results see Roels and Van Loon (2017, StatBel). Some of them, i.e. the ONS, suggest that web scraped data need special treatment while constructing price indices due to high frequency and volume of data, elevated levels of missing data and high product churn (Breton et al., 2016). Researchers give some recommendations on the choice of the index for particular product categories (Bhardwaj et al., 2017) but there are no unequivocal choices. Therefore, in our paper we do not use alternate price indices as in forecasting exercise the main goal is to mimic the National Statistical Office methodology in order to decrease forecast errors. Our approach is also in line with the work by Aparicio and Bertolotto (2017) as we think that if web scraped data can really improve forecasting, then this improvement should be achievable with simple approaches.

[3] Exercises carried out with different forms of filters showed that this use does not improve the results. We suspect that it may be the effect of a rigorous selection and classification of products, which removes atypical products.

3.4 Product selection and classification

In contrast to traditional survey-based collecting methods, web scraping collects information about all items available in stores in a fast and inexpensive way. Therefore, the eCPI database includes not only information on prices of most popular products, but the whole market offer. Theoretically, the possibility to include all available products should help to better track and predict price behavior. In practice, we find that the best forecasts are achieved when only products corresponding to those chosen by statistical offices are taken into account. Therefore, COICOP classification and weighting scheme is applied similarly to the methodology of Statistics Poland.

Forecasts of price dynamics based on all products, which do not use any weighting scheme, are highly inaccurate. There are significant differences between the unweighted online price index and the CPI in both short- and long term (see Fig. 6, Fig. 7). Moreover, the unweighted online price index does not show seasonal price changes correctly. The unsatisfactory performance of the unweighted online index originates from some basic goods like fruits, vegetables and other unprocessed or low-level processed products that are available online in one variant only. At the same time, many processed products are offered in different flavors, sizes and brands. Moreover, maintaining a wide offer in online stores is especially easy due to the lower cost of keeping products in stock and bigger area of activity than in traditional shops. Therefore, some varieties of products gain in importance in the unweighted price index while their price dynamics is not representative for the average household. To solve this problem we use CPI classification and a weighting scheme to reflect the importance of given products in household expenditures.

Fig. 6 Unweighted online price index, m-o-m.

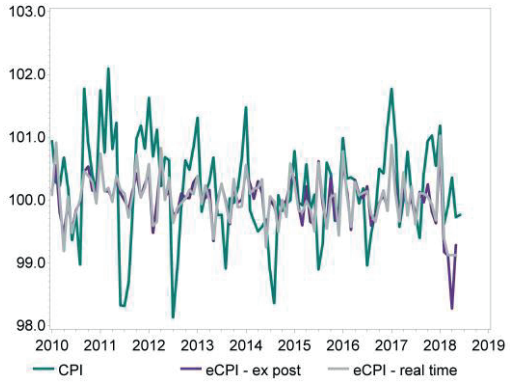
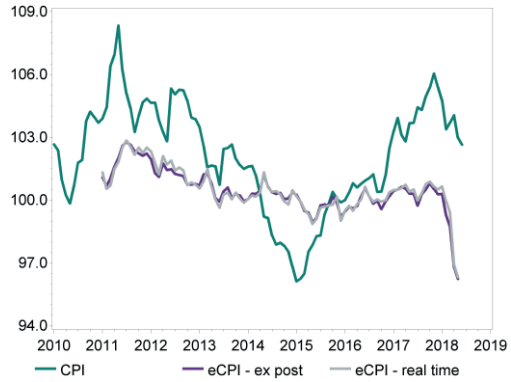


Fig. 7 Unweighted online price index, y-o-y.



We distinguish 3 main problems in the description of products classification that may lead to the deterioration of the quality of inflation aggregates. We denote the *classification error* as clearly and objectively inappropriate product assignment to the class (group) of products, e.g. apple classified to pears. The second one is *selection bias*, which results from clear mismatch of product varieties, even if the product itself is properly classified to the group. Statistics Poland (SP) collects only selected, representative products that are classified into COICOP groups. While our dataset contains various varieties of products including the one monitored by SP, selection bias occurs when we select a different one. It is also possible that our dataset does not include product variety exactly corresponding to the one used by SP. We call that problem the *product variety discrepancy*. These discrepancies result from sample size limitations (number of online retails stores) or differences between the online and offline product offer. They may also affect forecasting quality.

The Polish CPI of food and nonalcoholic beverages consists of 84 groups specified according to COICOP classification (see Eurostat, 2013). They are known as elementary groups as they are the most detailed, lowest-level of classification categories employed for CPI calculation. Every elementary group represents some part of individual household consumption. For each elementary group price indices are calculated and combined to higher level aggregates using weights based on the expenditures of households. Statistics Poland as well as other statistical offices collect

and use prices of selected products while calculating elementary groups indices. The chosen items are only significant ones, representative for purchases made by households and likely reflecting price movements similar to a wider range of goods and services. The product coverage limitation is mainly driven by the costs of survey data collection, which should be representative in terms of type and geographical location.

In our research we forecast each elementary group of food CPI and 10 main aggregates corresponding to 4-digit COICOP groups as well as the overall index of food and nonalcoholic beverages. We select products, which are possibly the most similar to those chosen by SP. To explain the importance of product selection we will use an example. Statistics Poland calculates the index of sugar based on the prices of one kilogram of white, regular sugar only. In online stores there are also other varieties of sugar available like cane, flavored, thick or powdered sugar. The dynamics of prices in online stores, which correspond to SP's representative goods match almost perfectly the official sugar CPI dynamics (see Fig. 8, Fig. 9). On the other hand, the online price index, which lacks variety selection, is significantly different from the official CPI.

Fig. 8 Sugar price index - one kilogram white, regular sugar, m-o-m.

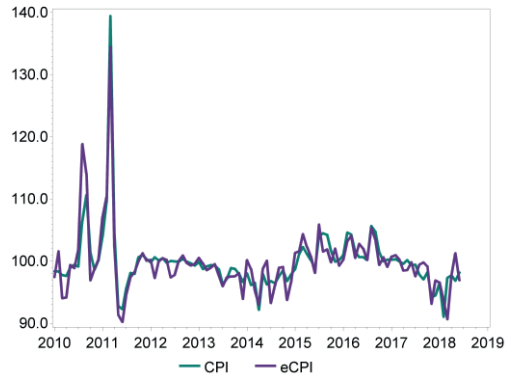
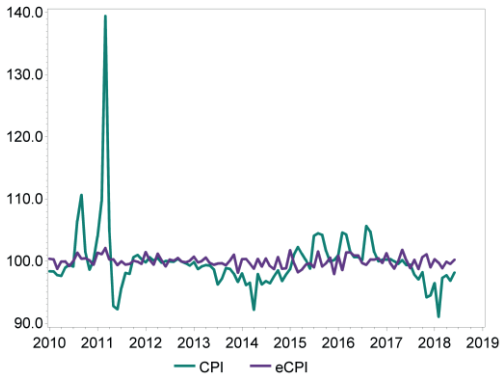


Fig. 9 Sugar price index - all kinds of sugar, m-o-m.



Such a big difference between the CPI and eCPI, when all varieties of sugar are included probably results from different pricing mechanisms of white, regular sugar and other less popular varieties. White sugar is a homogeneous product that is

difficult to distinguish and competes mainly by price. Considering these factors, the margins for white sugar are rather low and the price strongly depends on the costs of production, which in combination with high turnover implies high volatility of sugar prices. In contrast, other kinds of sugar are more distinctive, allowing their producers to have higher margins and are prone to small changes of production costs or other disruptions on the market (Fig. 10, Fig. 11).

Fig. 10 Histogram of sugar prices - one kilogram white, regular sugar, PLN.

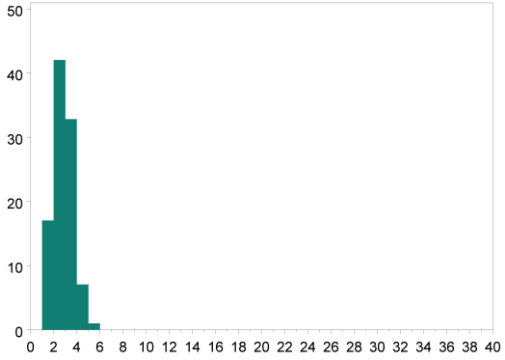
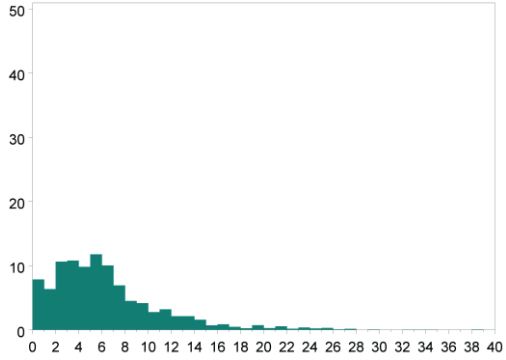


Fig. 11 Histogram of sugar prices - all kinds of sugar, PLN.



On the one hand, this result may suggest that the CPI does not represent inflationary processes accurately, because it omits many products available on the market. On the other hand, due to the overrepresentation of some products in online stores, other kinds of sugar probably have a significant influence on the price index. Quite a reasonable solution in this case would be to divide the sugars into two or more categories and weight them using the consumption structure for different types of sugar. However, due to the lack of such detailed information from the NSO, the optimal solution seems to be using prices of white sugar only because of its dominant position on the market. Moreover, the purpose of this research is to forecast the official CPI, so using solely products identical to SP representative goods seems to be the most reasonable solution. As some of the products available on the market are not included in the samples, eCPI figures may be biased in the same way as the CPI might be biased with respect to product coverage. Due to the very large number of

collected products a comprehensive analysis of price index sensitiveness to sample selection would require a separate article.

Due to the size of the eCPI data set, it is impossible to manually classify over 488 thousand products into one of the 84 elementary groups. Therefore, the process of allocating products to elementary groups is partially automated by using simple rules which analyze occurrences of specific words in the name product and category.

We do not rely on in-store categories which may be ill-suited to our classification. The product selection algorithm works as follows. In the first step product names and categories are unified. Text data are cleaned up, volume and weight measures converted to a common unit. If it is possible, we also apply some other rules to create product names (e.g. brand, product, volume). In the second step we select products if a specific string (usually the stem representing the key part of the name) was detected. It is a naive morphological method but it both limits the number of candidates, which helps to delete some of unrelated products, and is agnostic enough not to cut out word inflections excessively. In the third step we use similar rules to reject products which contain substrings indicating that the item should be classified to a different group or mismatched with respect to the SP's representative varieties.

To explain the selection stage we take sugar as an example again. At the beginning all product names, which contain the string 'sugar' pol. «cukier» are selected. However, the list still includes many products which do not fit into the 'sugar' group. We remove products that contain phrases like 'candies' «cukierki», 'sugar free' «bez cukru», 'reduced sugar' etc. – see Fig. 12. In the last step, we delete products which might belong to the elementary group, but are not similar enough to SP's representative product varieties (white sugar «biały cukier»). In the case of sugar, these are goods, which contain 'cane' «trzcinowy», 'brown' «brązowy», 'powdered' «puder» in the name string or are sold in smaller bags than one kilogram. Every month the selection is updated if a new product or variety appears and new rules are added if needed.

Fig. 12 White sugar selection based on prefiltered products as word clouds.

Note: Word clouds represent all of the words, which occurr in product names. Font size corresponds to frequency of the words. On the left, the word cloud of all product names which contain a string 'sugar' (pol. «cukier»). On the right, the word cloud of products after selection – white, regular sugar is the target.

So far in the eCPI project we have used manual and semi-automated classification methods of web stores' products to COICOP groups to maintain the highest quality of classification. It is mainly because we found fully automatic short text classification methods too erroneous at the time – unsupervised approach results in only 0.6 counted R^2 on average for food elementary groups. Automatic detection of the most significant phrase can be very challenging due to no specific order of noun and adjectives in Polish and the chaotic creation of product names by retailers. However, a decent classification tool would be a great help and we believe it is one of the next steps in the eCPI project development. We consider executing these tasks in the future on autopilot using word vector distances in a distributional framework (Word2Vec by Mikolov et al., 2013 or Fasttext by Grave et al., 2018) trained on Polish corpora. They look promising and do not depend on specific language grammar rules. A different approach to web scraped product classification is supervised classification with SVM (Breton et al., 2016) or other machine learning tools, e.g. neural networks.

Our primary online index for food aggregate, constructed with the methodology explained in this chapter is called eCPI, although eCPI itself is produced and evaluated in two variants. The first one (the ex-post variant) uses all available data

collected during the given month, so it may be perceived as an alternative method of calculating inflation. The second one (the real-time variant) uses only data available mid-month, just after the official monthly CPI releases in Poland. Therefore, we use online food prices from roughly two–three weeks in the current month. The dynamics of the eCPI and eCPI real-time fairly resembles the official CPI dynamics, although there are discrepancies in some periods – see Fig. 13 and Fig. 14.

Fig. 13 Food and non-alcoholic beverages, m-o-m.

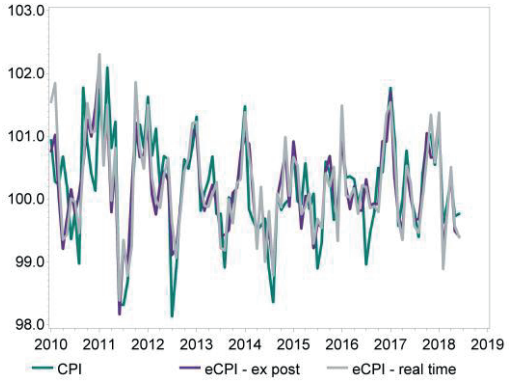
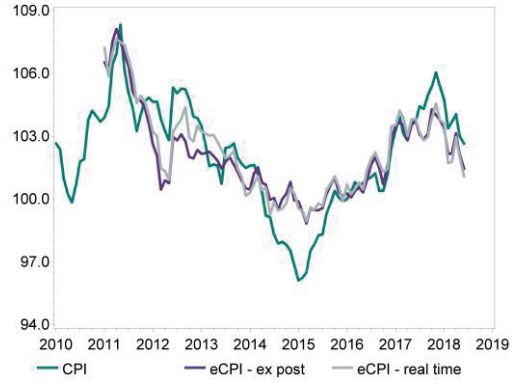


Fig. 14 Food and non-alcoholic beverages, y-o-y.



Clearly web scraped prices differ in many aspects from official prices. Firstly, products collected by an NSO may and certainly do differ from products scraped with our routines. Secondly, while an NSO collects products rather selectively but covers the domestic shops' sample better, our routine does the opposite. It scrapes all of the available prices from a small number of shops, which undermines the representativeness of the data for purposes of official statistics. Thirdly, our frequency of price collecting differs extremely from an NSO's. As higher frequency typically lowers the risk of recording outliers, the frequency differences may vary among countries resulting in fewer (or more) gains in forecasting. For example Statistics Poland collects the prices of fruit and vegetables twice per month, while, as Aparicio and Bertolotto (2017) report, some other NSOs collect only once (or even less frequently). Moreover, in some countries there are different approaches to the missing prices, like price imputations for missing products for no more than 7 days (see Aparicio and Bertolotto, 2017), or different price recording frequency in specific

regions. Therefore a comparison of the online index fit among countries and studies is rather difficult.

3.5 Forecasting experiment scheme

Using online prices gives a unique opportunity to apply higher-frequency data to nowcast the CPI, as well as analyze inflation developments even before the actual publication of official statistics. Analyzing data from online stores is possible with very small delays, while inflation is published generally on a monthly basis.

We use a real-time dataset to forecast the m-o-m index of food inflation in Poland. The verification period extends from January 2014 to June 2018, while the initial estimation period, being limited by web scraped data availability, starts with January 2010 and ends in December 2013.

Our intuition is that fast-arriving data in the current month should mostly improve nowcasting. According to Cavallo and Rigobon (2016) price levels calculated from online data may deviate significantly from the official ones, while their price dynamics generally behaves similarly and quickly reacts to aggregate shocks. However, the common movements may still differ in scale, hence they may deviate from the official data even in the short term. Therefore, we check the forecasting gains of web scraped data included in simple linear regression models. In our article we estimate simple autoregressive distributed lag models – henceforth ADL – according to eq. (1) for each of the 84 elementary groups and additional 10 food subaggregates. In total we estimate 72 specifications by including or excluding the AR part (zero restriction on β_i parameters), or deterministic factors (zeros on δ_i) and experimenting with lag orders.

$$\pi_t = \sum_{i=1}^{N} \beta_i \pi_{t-i} + \sum_{i=0}^{M} \gamma_i x_{t-i} + \sum_{i=1}^{12} \delta_i d_i + \varepsilon_t, \qquad \varepsilon_t \sim NIID \tag{1}$$

π_t – a monthly, non-seasonally adjusted official price index of elementary COICOP group or food inflation subaggregate,

x_t – a non-seasonally adjusted eCPI index (based on web-scraped, daily, online prices), aggregated into monthly frequency,

N and M are lag orders which vary, $N = 1, 2, 3, 6, 12$ and $M = 1, 2, 3, 6, 12$,

d_i – seasonal deterministic factors,

ε_t – a normal, independent and identically distributed error term.

We call the models specified by equation (1), where $\gamma_i \neq 0$ for at least one i, the eCPI-in-ADL. To distinguish the eCPI series itself, which provides the current value x_0 as a nowcast from model-based approaches we refer to the eCPI as the raw index or raw eCPI.

In order to evaluate the quality of forecasts of the eCPI and benchmarks we calculate out-of-sample: the root mean square forecast error (RMSFE), the mean absolute forecast error (MAFE) and the mean forecast error (MFE). Forecast error measures are evaluated on 24-month windows. In addition, we compare the accuracy of the forecasts using HLN Diebold-Mariano (1995) and Giacomini-White (2006) tests. In the results section, however, we report only the former since both procedures give similar results.

To reduce model selection bias we also use a linear combination of forecasts with equal weights and the weights inversely proportional to RMSFEs. As the simple mean approach provides small gains in accuracy, we proceed with weights inversely proportional to RMSFEs and we report results for this variant only. We realize that in some product groups the online data may fit better than in others. Forecasting price dynamics in low-level groups of products like elementary groups offer potential benefits from aggregation, selection or combination – we assess 84 x 72

forecasts in total. Specification selection applies to every elementary group and then the selected forecasts are aggregated according to SP methodology into our target variable, food inflation.

We examine several benchmark models that do not include online data information. In a pseudo real-time experiment we evaluate the random walk $\pi_t = \pi_{t-1} + \varepsilon_t$ (RW), the seasonal random trend model $\pi_t = \pi_{t-1} + \pi_{t-12} - \pi_{t-13} + \varepsilon_t$ (SRW), the ARMA ($\pi_t = \alpha + \sum_{i=1}^{p} \beta_i \pi_{t-i} + \sum_{i=1}^{Q} \theta_i \varepsilon_{t-i} + \varepsilon_t$) of lag order (P,Q) as well as the SARMA selected by the Akaike criterion, the ARMA selected in (pseudo) real time by the 24 months RMSFE (only the one, which has the lowest value) and a linear combination of ARMA models (weighted average of forecasts) using the weights inversely proportional to their RMSFEs. As a baseline we choose the ARMA model estimated for the CPI food aggregate with the lowest forecast error across specifications, based on a 24 months window RMSFE in pseudo real-time.

4 Results

We find that disaggregated random walks perform poorly, being beaten severely by the baseline (ARMA aggregate). The utilizing of disaggregated information in ARMA models selected in real time by the RMSFE improves the accuracy by 7% with respect to the baseline (see Tab. 3). Forecast combination based on RMSFE weights instead of picking up the best ARMA specification reduces the error a little bit further. It seems that benchmarks cannot go over a 10% RMSFE reduction.

Tab. 3 RMSFE of m-o-m food inflation nowcasts, Jan 2014 - June 2018.

Approach	Description	RMSFE	Relative
benchmark (baseline)	ARMA (aggregate, best)	**0.57**	**1.00**
other benchmarks (without eCPI)	RW	0.72	1.26
	Seasonal RW	0.68	1.18
	ARMA AIC	0.58	1.01
	ARMA (best)	0.53	0.93
	ARMA forecasts combination	0.52	0.90
eCPI-based forecasts	**eCPI (real-time)**	**0.47**	**0.82**
	eCPI ex post	0.44	0.78
	eCPI-in-ADL (best)	0.41	0.73
	eCPI-in-ADL forecasts combination	**0.40**	**0.70**
	Best selected from ARMA, eCPI and eCPI-in-ADL	0.40	0.70
	--- forecast combination	0.43	0.75

Note: all of the approaches are based on the lowest-level COICOP group forecasts (elementary groups) except for the baseline, which is estimated on the total food inflation (CPIF) aggregate.

Tab. 4 Diebold-Mariano test results for m-o-m food inflation nowcasts, based on RMSFE.

Model tested	against the model (baseline)	p-val
eCPI real-time	Seasonal RW	0.00
	ARMA (aggregate, best)	0.06
	ARMA (best)	0.14
	eCPI ex post	0.83
forecast combination of eCPI and eCPI-in-ADL	ARMA (aggregate, best)	0.00
	ARMA (best)	0.00
	ARMA forecasts combination	0.00
	eCPI (real-time)	0.05
	eCPI ex post	0.23
	Best selected from eCPI and eCPI-in-ADL	0.27
Best selected from ARMA, eCPI and eCPI-in-ADL	ARMA forecasts combination	0.00
	eCPI (real-time)	0.10
	eCPI ex post	0.26
	forecast combination of eCPI and eCPI-in-ADL	0.47

Note: p-val indicates a p-value for the hypothesis pair: H_0: $\sigma_{left\ hand\ side\ model} > \sigma_{right\ hand\ side\ model}$, H_1: $\sim H_0$, where by σ we denote the RMSFE calculated in the period Jan2014:Jun2018.

Raw eCPI in real time reduces forecast errors by almost 20% with respect to the baseline, which we find significant at the edge (Diebold-Mariano test p-value of 0.06). However, when we consider both the raw eCPI and eCPI-in-ADL and select the best one in real time, we obtain a 27% improvement over the baseline. Forecast combination instead of selection of only one of these models results in further improvement, by 30%, of the RMSFE relative to the baseline. The combination of eCPI and eCPI-in-ADL clearly beats ARMA models (p-value of 0.00) and likely real-time raw eCPI (p-value of 0.05). Although in terms of the Diebold-Mariano test there is not much evidence that the forecast combination performs significantly better than the ex-post calculated eCPI (which consists of full month information) or simple selection (see Tab. 4).

Finally, we test a forecast selection from a wide range of approaches: the raw eCPI, the eCPI-in-ADL and the ARMA evaluating, as previously, their 24 months RMSFE. While this approach is clearly superior to the baseline, it is hardly better than eCPI real-time (D-M p-value of 0.10) and we find no real benefits of this combination over the forecast combination of the eCPI and the eCPI-in-ADL (Tab. 4). To sum up, the results of this small 'horse race' suggest online prices help to improve food inflation nowcasts, however, it is more plausible to incorporate this information into a simple linear regression than to use raw online indices. This observation may result from mitigating short-term deviations of online indices from the offline ones. In short-term we can benefit from some additional adjustments even in such a simple model as the ADL.

We find the disaggregated forecast of food inflation superior to aggregate forecast, or in other words a greater RMSFE reduction due to adopting web scraped data in the disaggregated approach than in the aggregated one. Powell et al. (2018) mention product classification errors as a possible explanation of worse forecasting abilities. We find that this argument mostly explains the case, and we believe the difference in product coverage with respect to the NSO basket plays the main role. In addition, the web scraped data sample is limited, which may cause some noise or bias, thus eCPI data inclusion in the ADL model further improves the results. Our pseudo real-time experiment suggests that such a mixed disaggregated approach provides a superior nowcasting performance.

Detailed results in disaggregated forecasts present a rather mixed picture of performance improvement. While the raw eCPI in real time reduces forecast errors by 18% with respect to the best ARMA model for the food inflation aggregate, it reduces forecast errors only in 17% of elementary groups (mainly unprocessed food, fruits and vegetables in particular, see Fig. 15 and Tab. 5 - Tab. 8 in Appendix). Differences appear mainly for the monthly rate of growth, while the fixed base indices reveal similar trends and reflect well the long-term trend of the CPI. If we use

combined eCPI-in-ADL models and compare them to a combination of ARMA models we get more accurate forecasts in 79% of elementary groups (see Fig. 16).

Fig. 15 eCPI and the best ARMA by elementary group.

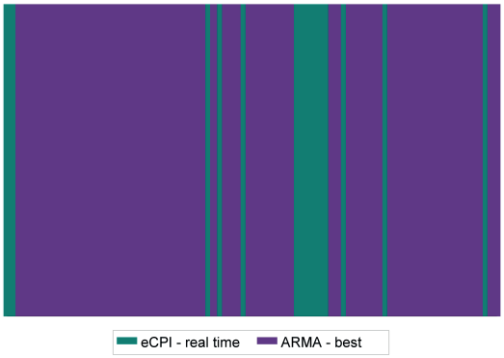
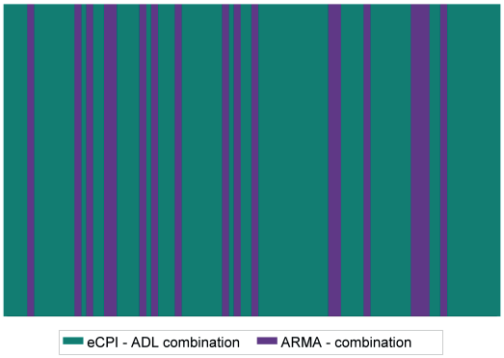


Fig. 16 eCPI-in-ADL combined and the combination of ARMA by elementary group.

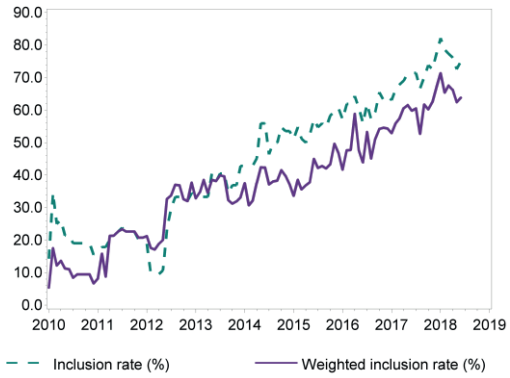


Note: Figures show which approach gives a more accurate forecast in terms of RSMFE for each elementary group. The consecutive COICOP groups are on the x-axis, while y-axis is the null axis. The raw, real-time eCPI has a lower RMSFE than ARMA in only 17% of elementary groups. However, the eCPI-in-ADL combination is more accurate than the ARMA combination in 79% of elementary groups.

Our findings indicate that forecasts based on data from online stores are especially accurate for those groups with high price volatility (mainly unprocessed food, fruits and vegetables in particular). We attribute this phenomenon to the fact that in the presence of a common shock on a competitive market retailers are forced to change prices in a coordinated manner. When there are moderate price swings, individual differences between retailers become more important (e.g. different suppliers, contracts, pricing policy), which results in big differences between the CPI and the forecasts from online stores.

In our opinion, ADL models and their combination allow us to eliminate bias and extract forecasting properties of eCPI data that when taken alone may exhibit short-term deviations from offline prices.

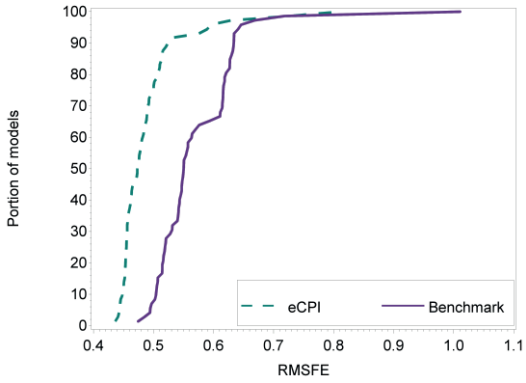
Fig. 17 Predictive inclusion rate of eCPI data as percentage of product groups



Note: The figure presents a percentage of elementary groups for which forecasts with the eCPI are more accurate than benchmarks in terms of RSMFE calculated on a 24-months window. The weighted inclusion rate accounts for the weights system for the CPI.

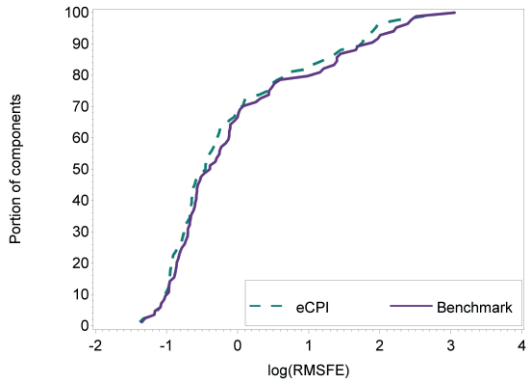
Online data quality is likely to increase over time. We denote an eCPI inclusion rate as the proportion (%) of elementary groups in which the model that includes the eCPI features a lower RMSFE in a particular month, so it is preferred in a real-time forecasting experiment. This measure is mostly increasing over time (see Fig. 17), in both views, ordinary and weighted with CPI basket weights. Obviously the rise in the very beginning of the sample may be linked to low online data availability relative to the official CPI time series. As the number of web stores and unique products collected increases over time, the inclusion rate reaches 75% in mid 2018.

Fig. 18 Cumulative distribution of RMSFE across model specifications.



Note: The figure presents the RMSFE of the models pairs as follow: including the online eCPI series and excluding it (benchmarks). Forecast errors are sorted by models. The solid line depicts benchmarks (offline prices only), and the dashed one the eCPI (including online data).

Fig. 19 Cumulative distribution of RMSFE across food inflation components.



Note: The figure presents the RMSFE of the best selected model from ADL and ARMA specifications for each pair: including the online eCPI series and excluding it (benchmark) for each elementary group. Forecast errors are sorted by food components. The solid line depicts benchmarks (offline prices only), and the dashed one the eCPI (including online data).

In general, forecast error reduction present in the best specification might not be sufficiently convincing for applied economists to start using such data. We report a forecast error figure similar to the one in Aparicio and Bertolotto (2017), as we find it useful in providing a quick summary of online price benefits to the potential forecaster. Fig. 18 compares forecast errors of specifications including eCPI to those without online price information (benchmarks) in a cross-section of 72 corresponding specifications. It does not matter which model specification the potential forecaster picks up, he can clearly outperform it by including eCPI. Additionally, we report a similar figure of the RMSFE, but as a function of food components (see Fig. 19), which confirms that for most of the components eCPI improves forecasts in terms of the RMSFE. Both views clearly confirm that fast-arriving online data are beneficial in the short-term forecasting of food inflation.

5 Conclusions

In the recent years we have observed an impressive development of the e-grocery market in Poland. New methods of data collection like web scraping offer an opportunity to collect and utilize online prices in the inflation forecasting process. In this paper we assess the ability of online prices to improve food inflation nowcasting using our own data warehouse based on web scraped data collected from 2009 till mid-2018. We perform pseudo real-time forecasting experiments, both for the food inflation aggregate and its 84 subaggregates.

Our main finding is that the most successful approach for incorporating online prices to produce inflation forecasts consists of 3 key elements: 1) proper product selection and classification, 2) aggregating components with official expenditure weights in line with statistical office methodology and 3) combining simple models including online data for each group. Online price data improve food inflation nowcasts in Poland significantly and outperform the benchmark models.

We find product selection and classification as well as proper result aggregation a very important issue in applying online data into the forecasting process. Considering unit product level, we realize that in online stores product coverage is different to the products collected by Statistics Poland. Web scraping allows collecting information about all available items in stores in a fast and inexpensive way whereas official CPI includes only selected products. We found that in practice the best forecasts are produced when only products similar to these selected by Statistics Poland are used and CPI official weighting scheme is applied.

The raw real-time eCPI reduces forecast errors by almost 11% with respect to the best ARMA models. We find this result favorable as it allows regular forecasters to improve food inflation nowcasting in only two months after the start of collecting online data. The web scraped price index alone enhances nowcasts particularly well in the most volatile groups of goods, which are the most difficult to forecast by

standard models. We attribute this phenomenon to the fact that in the presence of a common shock on a competitive market retailers are forced to change prices in a coordinated manner. In the case of products being subject to moderate price swings results are less favorable in the short term due to the unsynchronized process of price change. However, in the long term discrepancies between eCPI and CPI are reduced.

We find that incorporating eCPI data into simple, linear regression models is a superior approach to inflation nowcasting as it improves performance of forecasts by eliminating bias and extracting forecasting properties of eCPI data series, which alone may deviate from official data in the short term. We show that the eCPI in the role of regressor reduces forecast errors in most of the groups and models. Further gains can be achieved by combining forecasts from different models, however, the difference is not statistically significant.

Our future research will focus on forecasting the total CPI with web scraped data. In fact this objective may not be far from our current research progress as food carries a significant weight in the CPI weighting system and in most studies the easily obtainable weekly offline fuel prices are used providing satisfactory results. We expect that mastering automatic product classification could be beneficial in larger scale calculation of price indices.

It remains an open question whether forecasting gains from online prices span beyond the current month. Our preliminary results suggest this hypothesis is true, which is in line with the statement by Faust and Wright (2013) – in order to produce better long-term forecasts, one should firstly improve the one-step ahead forecast. This area certainly needs further research.

6 References

Aghajanyan G., Baghdasaryan T. and Lazyan G. (2017). The use of Big Data in Central Bank of Armenia, IFC-Bank Indonesia Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference, Bali, Indonesia.

Aparicio, D. and Bertolotto, M. (2017). Forecasting Inflation with Online Prices (June 1, 2017), http://dx.doi.org/10.2139/ssrn.2740600

Bermingham, C. and D'Agostino, A. (2011). Understanding and Forecasting Aggregate and Disaggregate Price Dynamics, ECB Working Paper No. 1365.

Bertolotto M., Cavallo A. and Rigobon R. (2014). Using Online Prices to Anticipate Official CPI Inflation, UTokyo Price Project Working Paper Series 031, University of Tokyo, Graduate School of Economics.

Bhardwaj H., Flower T., Lee P., Mayhew M. (2017). Research indices using web scraped price data, Office for National Statistics.

Breton R., Flower T., Mayhew M., Metcalfe E., Milliken N., Payne C., Smith T., Winton J., Woods A. (2016). Research indices using web scraped data, Office for National Statistics.

Buono D., Mazzi G. L., Kapetanios G., Marcellino M. and Papailias F. (2017). Big data types for macroeconomic nowcasting, in: Eurona (EUrostat Review On National Accounts and macroeconomic indicators), 1/2017.

Cavallo, A. (2013). Online and official price indexes: Measuring Argentina's inflation, Journal of Monetary Economics, Elsevier BV, 2013, Vol. 60(2), pp. 152-165.

Cavallo, A. (2017). Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers, American Economic Review, American Economic Association, 2017, Vol. 107(1), pp. 283-303.

Cavallo, A. (2018). Scraped data and sticky prices. 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, American Economic Association, 2016, Vol. 30(2), pp. 151-178.

COICOP five-digit structure and explanatory notes. (2013). Unit B5 "Management of Statistical Data and Metadata", Eurostat.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics Vol. 13 (3), pp. 253–63.

E-grocery w Polsce - zakupy spożywcze online, report, Mobile Institute, 2017. https://www.ecommercepolska.pl/files/4415/1775/0535/E-grocery_w_Polsce_Zakupy_spozywcze_online_raport.pdf

Faust, J. and Wright, J. H. (2013). Forecasting Inflation, Chapter 1, pp. 2-56, Elsevier.

Giacomini, R. and White, H. (2006). Tests of Conditional Predictive Ability, Econometrica, Vol. 74, No. 6, pp. 1545-1578.

Grave E., Bojanowski P., Gupta P., Joulin A. and Mikolov T. (2018). Learning Word Vectors for 157 Languages, Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018), source: https://github.com/facebookresearch/fastText/blob/master/docs/crawl-vectors.md

Greenaway, M. (2018). ONS Web-scraping policy, Office for National Statistics.

Hull I., Löf M., Tibblin M. (2017). Price information collected online and short-term inflation forecasts, IFC-Bank Indonesia Satellite Seminar on "Big Data" at the ISI Regional Statistics Conference, Bali, Indonesia.

Huwiler, M. and Kaufmann, D. (2013). Combining disaggregate forecasts for inflation: The SNB's ARIMA model, No 2013-07, Economic Studies, Swiss National Bank.

Lunnemann, P. and Wintr, L. (2006). Are Internet Prices Sticky? ECB Working Paper No. 645 (June 2006).

Melis K., Campo K., Lamey L. and Breugelmans E. (2016). A Bigger Slice of the Multichannel Grocery Pie: When Does Consumers' Online Channel Use Expand Retailers' Share of Wallet? Journal of Retailing, Elsevier BV, 2016, Vol. 92(3), pp. 268-286.

Mikolov T., Chen K., Corrado G., Dean J. (2013). Efficient Estimation of Word Representations in Vector Space, in Proceedings of Workshop at ICLR.

Nielsen Grocery Universe 2017. (2017). Results of the 55th Inventory of Retail Grocery in Belgium, The Nielsen Company.

Powell B., Nason G. , Elliott D., Mayhew M., Davies J. and Winton J. (2018). Tracking and modelling prices using web-scraped price microdata: towards automated daily consumer price index forecasting. J. R. Stat. Soc. A, 181: 737-756.

Radzikowski, B. and Smietanka, A. (2016). Online CASE CPI, Proceedings of the 1st International Conference on Advanced Research Methods and Analytics, Universitat Politècnica València.

Roels, D. and Van Loon, K. (2017). Le webscraping, la collecte et le traitement de données en ligne pour l'indice des prix à la consommation, slides, StatBel.

Swier, N. (2014). How should web scraping be organised for official statistics, presentation slides, https://ec.europa.eu/eurostat/cros/system/files/Presentation%20S6AP3.pdf

Szafranek, K. (2017). Bagged artificial neural networks in forecasting inflation: An extensive comparison with current modelling frameworks, NBP Working Papers 262, Narodowy Bank Polski.

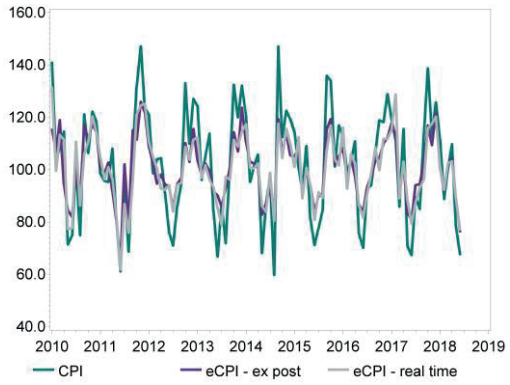
Szafranek, K. and Hałka, A. (2017). Determinants of low inflation in an emerging, small open economy. A comparison of aggregated and disaggregated approaches, NBP Working Papers 267, Narodowy Bank Polski.

The Digitalization Of Food: Grocery Retail And Food Service (2016). Fung Global Retail & Technology report, July 14, 2016.

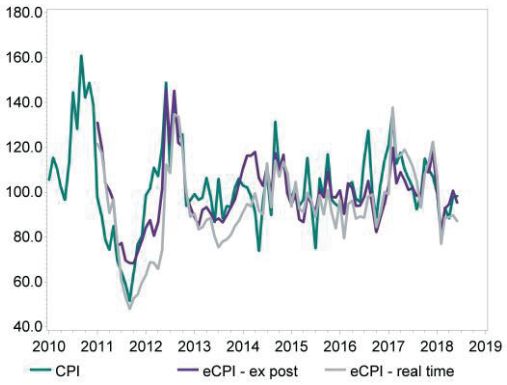
7 Appendix

Fig. 20 (Appendix) eCPI performance for unprocessed food (selected groups).

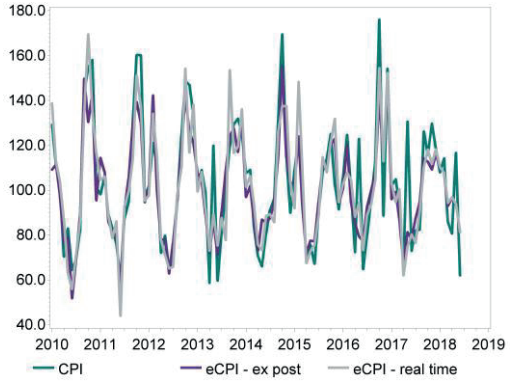
Tomatoes, m-o-m.



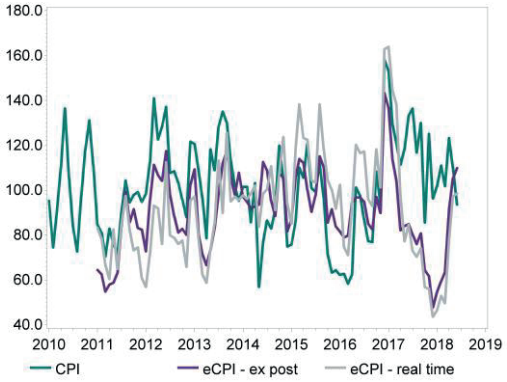
Tomatoes, y-o-y.



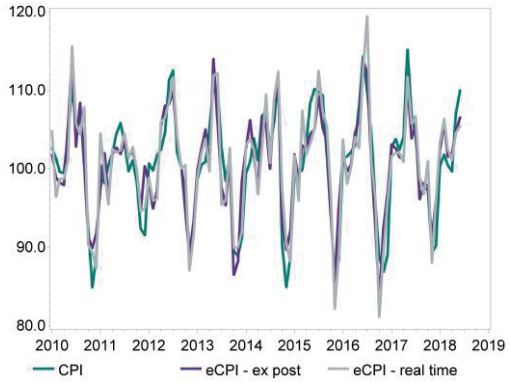
Cucumbers, m-o-m.



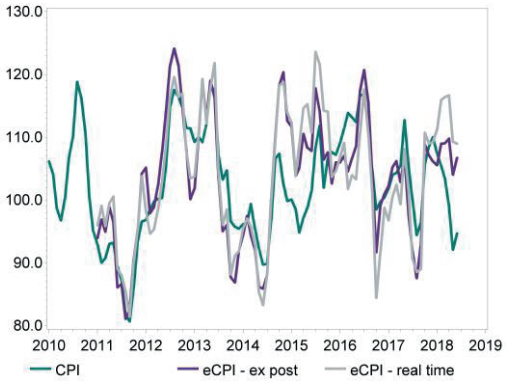
Cucumbers, y-o-y.



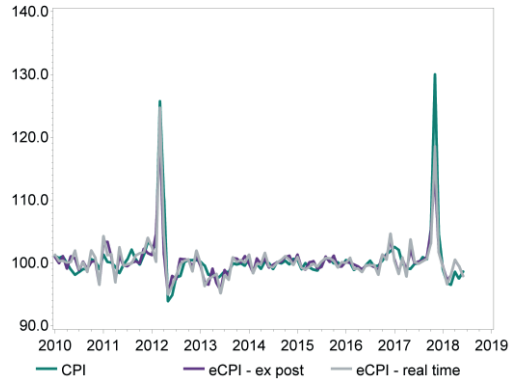
Citrus fruits, m-o-m.



Citrus fruits, y-o-y.



Eggs, m-o-m.



Eggs, y-o-y.

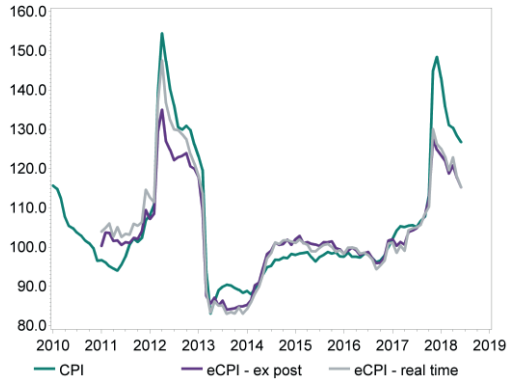
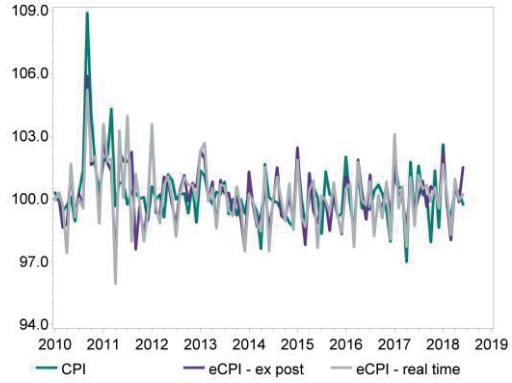
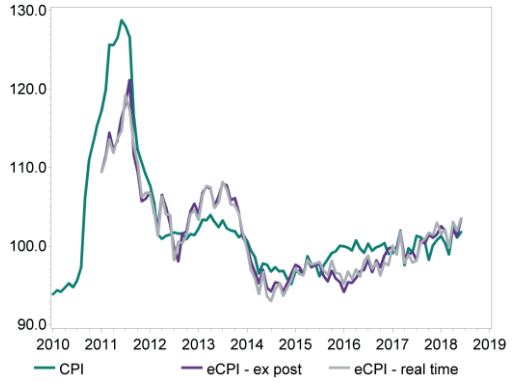


Fig. 21 (Appendix) eCPI performance for processed food (selected groups).

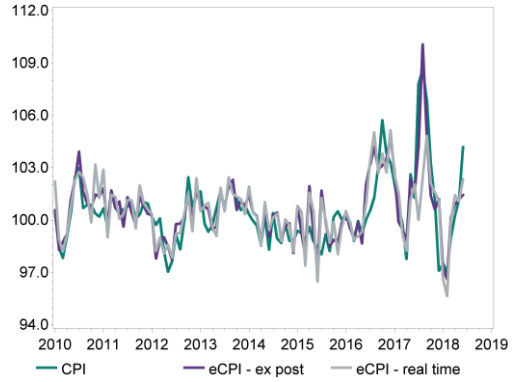
Wheat flour, m-o-m.



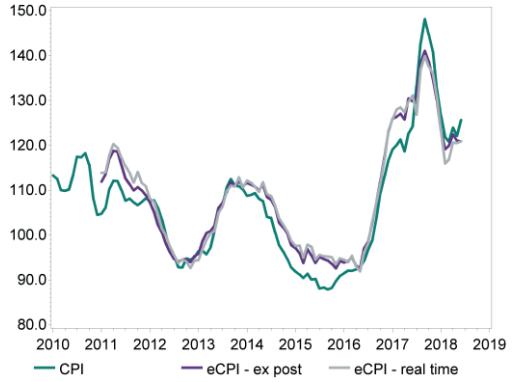
Wheat flour, y-o-y.



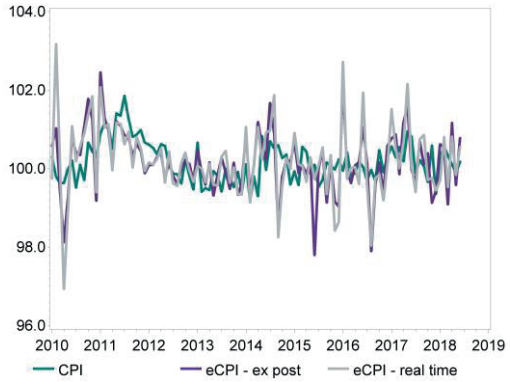
Butter, m-o-m.



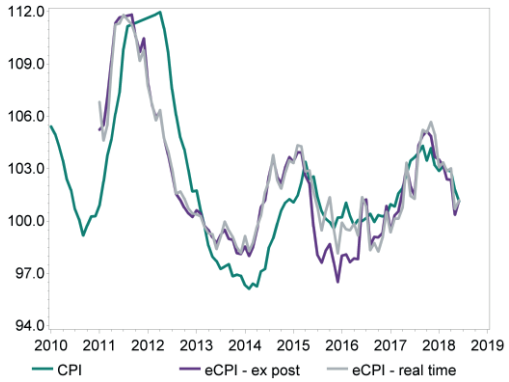
Butter, y-o-y.



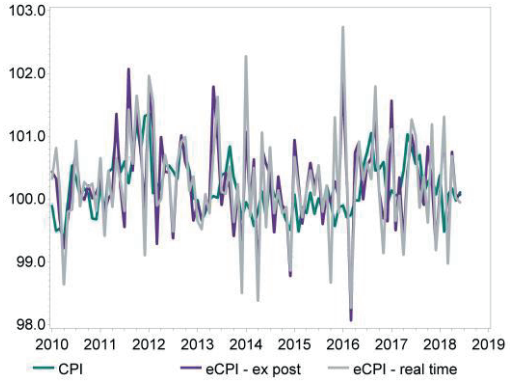
Coffee, m-o-m.



Coffee, y-o-y.



Dried, salted or smoked meat, m-o-m.



Dried, salted or smoked meat, y-o-y.

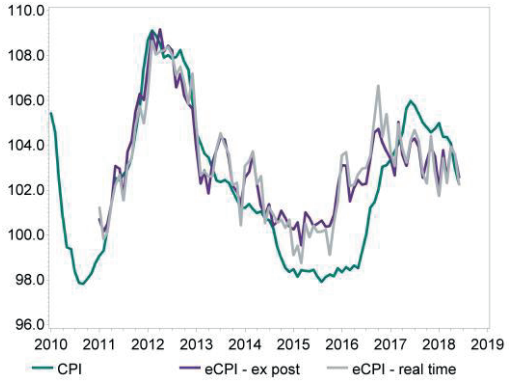
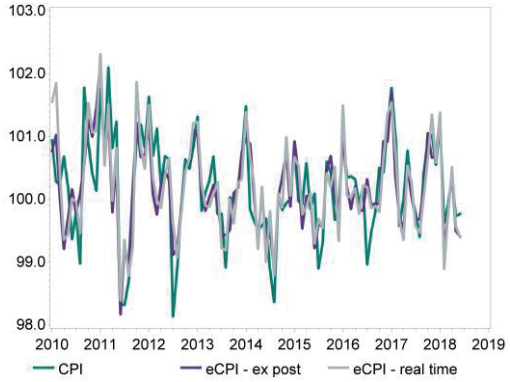


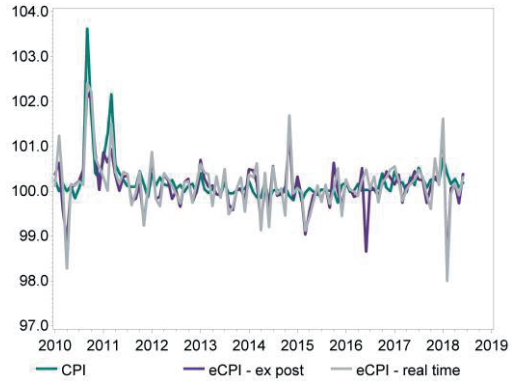
Fig. 22 (Appendix) eCPI performance for aggregates.

Food and non-alcoholic beverages, m-o-m.

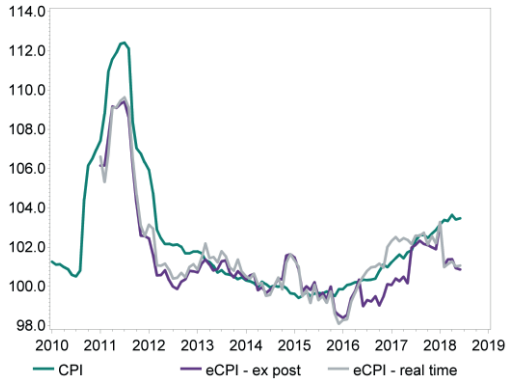


Food and non-alcoholic beverages, y-o-y.

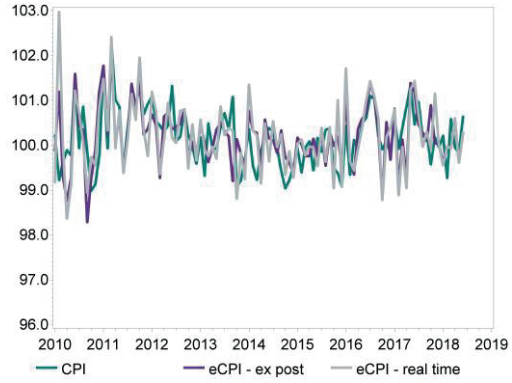
Bread and cereals, m-o-m.



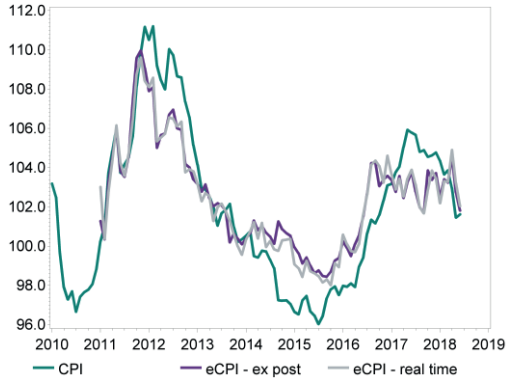
Bread and cereals, y-o-y.



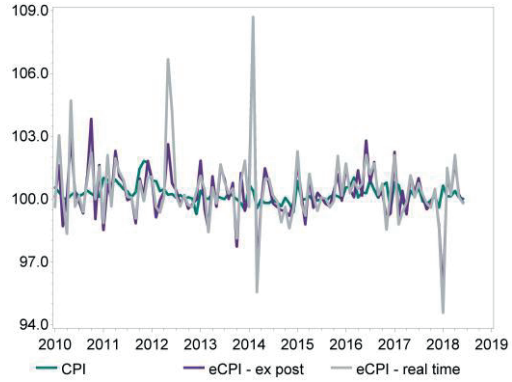
Meat, m-o-m.



Meat, y-o-y.

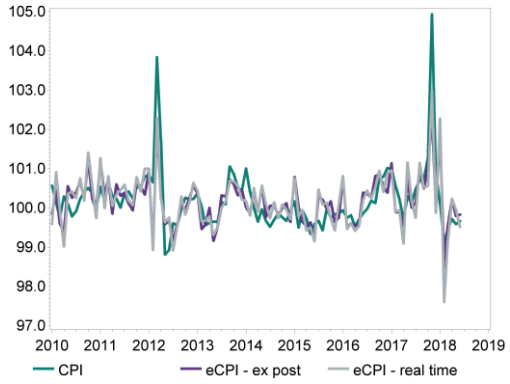


Fish and seafood, m-o-m.

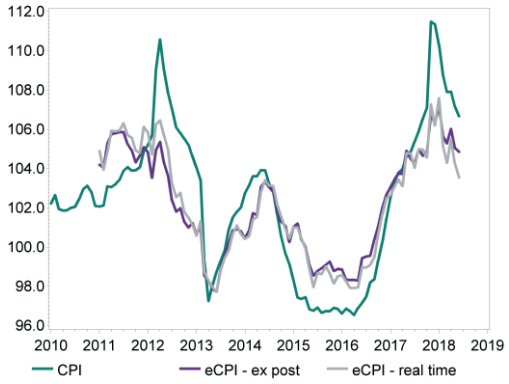


Fish and seafood, y-o-y.

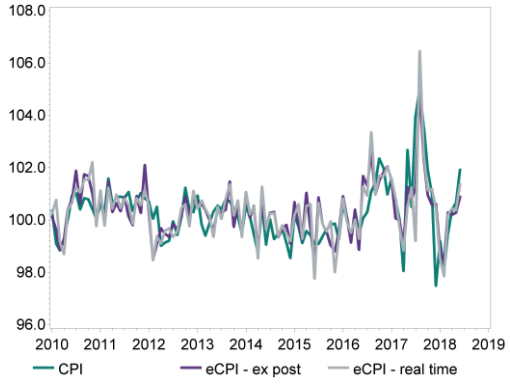
Milk, cheese and eggs, m-o-m.



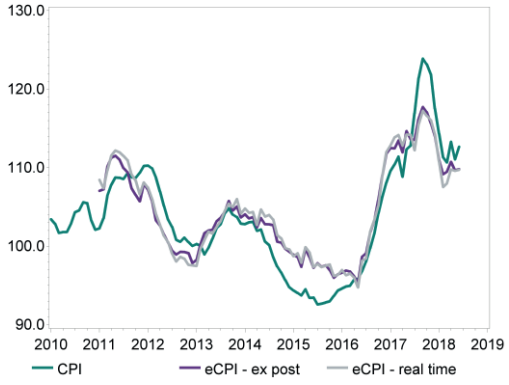
Milk, cheese and eggs, y-o-y.



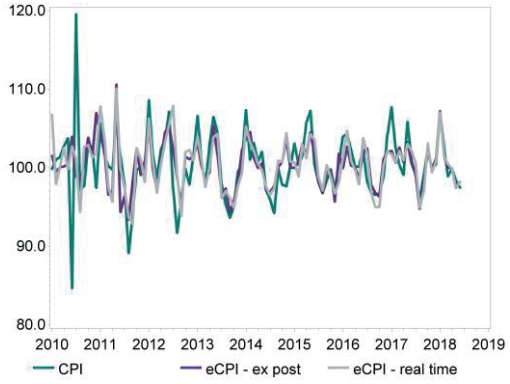
Oils and fats, m-o-m.



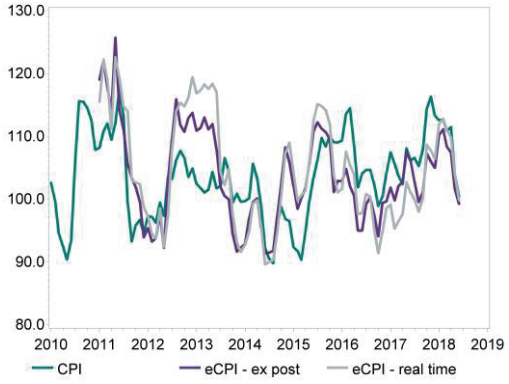
Oils and fats, y-o-y.



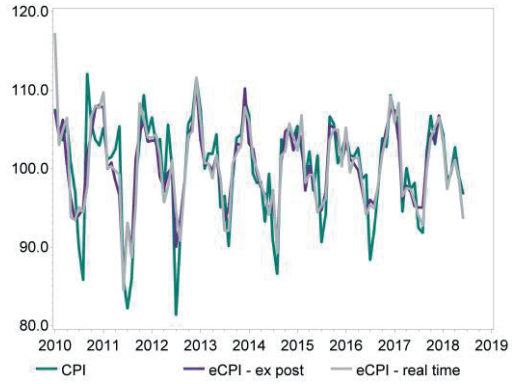
Fruits, m-o-m.



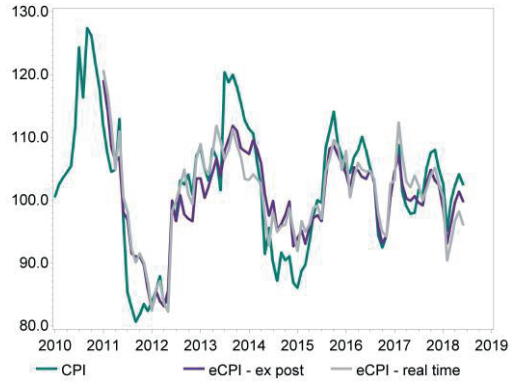
Fruits, y-o-y.



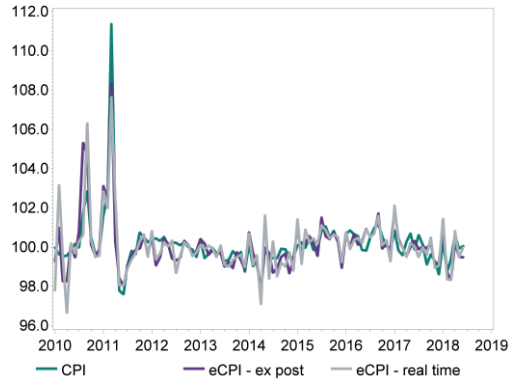
Vegetables, m-o-m.



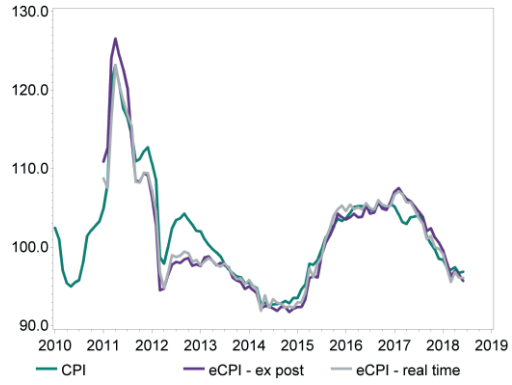
Vegetables, y-o-y.



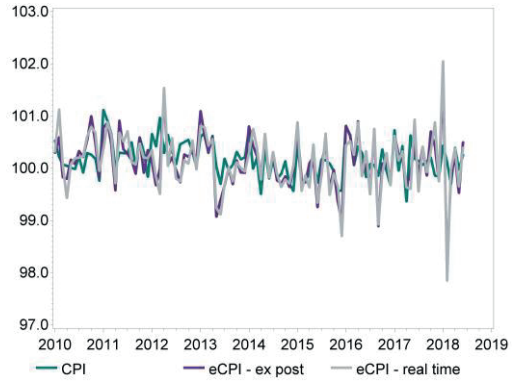
Sugar, jam, honey, chocolate and others, m-o-m.



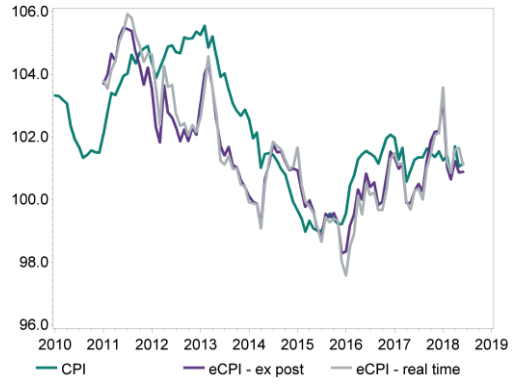
Sugar, jam, honey, chocolate and others, y-o-y.



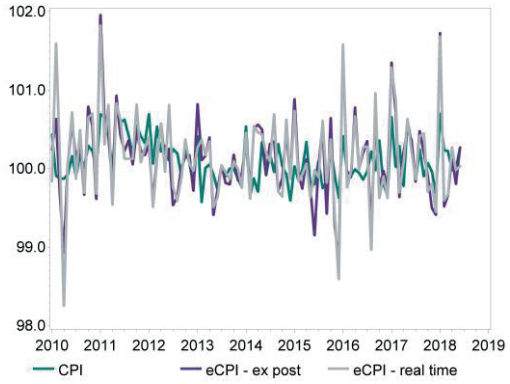
Food products n.e.c., m-o-m.



Food products n.e.c., y-o-y.



Non-alcoholic beverages, m-o-m.



Non-alcoholic beverages, y-o-y.

Tab. 5 (Appendix) RMSFE of eCPI and benchmarks nowcasts, Jan 2014 - Jun 2018.

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Food and non-alcoholic beverages	0.443	0.471	0.420	0.404	0.434	0.530	0.521	0.520	0.577	0.718
Bread and cereals	0.500	0.552	0.236	0.323	0.338	0.176	0.223	0.234	0.211	0.224
Rice	0.852	1.013	0.763	0.698	0.696	1.018	0.999	1.001	1.041	1.163
Wheat flours	1.070	0.987	1.050	0.981	0.989	1.342	1.289	1.291	1.455	1.966
Other flours	1.312	1.356	0.956	0.887	0.887	0.933	0.957	0.967	0.976	1.323
Groats and grains	0.669	0.836	0.429	0.388	0.388	0.398	0.410	0.411	0.428	0.489
Bread	0.819	0.805	0.399	0.560	0.589	0.253	0.313	0.340	0.298	0.263
Other bakery products	0.608	0.680	0.296	0.255	0.257	0.255	0.260	0.261	0.268	0.353
Pizza and quiche	0.773	1.035	0.358	0.367	0.367	0.389	0.378	0.383	0.386	0.621
Pasta products and couscous	0.738	0.994	0.485	0.471	0.473	0.498	0.517	0.519	0.527	0.726
Breakfast cereals	0.476	0.480	0.299	0.317	0.317	0.345	0.355	0.356	0.365	0.439
Other cereal products	0.985	1.156	0.432	0.515	0.518	0.607	0.574	0.574	0.587	0.682
Meat	0.565	0.650	0.457	0.439	0.442	0.421	0.374	0.373	0.454	0.585
Beef	1.708	1.804	0.444	0.385	0.391	0.499	0.496	0.502	0.488	0.456
Veal	1.960	2.381	0.272	0.267	0.282	0.303	0.338	0.343	0.332	0.312
Pork	1.545	1.628	1.170	1.087	1.086	1.072	1.011	1.010	1.090	1.362
Lamb and goat	2.508	2.954	1.439	0.742	0.771	0.881	0.889	0.889	0.881	1.213
Chickens	1.880	1.859	1.603	1.625	1.634	1.836	1.553	1.565	1.809	2.790
Other poultry	3.568	3.583	0.871	0.785	0.790	0.937	0.898	0.900	0.971	1.082
Other meats	2.560	3.226	2.199	1.369	1.356	1.653	1.622	1.626	1.702	2.135
Edible offal	1.573	1.939	0.479	0.433	0.441	0.415	0.419	0.419	0.599	0.443
Dried, salted or smoked meat other than poultry	0.731	0.905	0.359	0.324	0.329	0.299	0.272	0.270	0.342	0.344
Dried, salted or smoked poultry meat	1.289	1.667	0.364	0.340	0.343	0.433	0.382	0.382	0.388	0.468
Mixed ground meat	3.134	4.219	0.810	0.775	0.784	0.847	0.796	0.796	0.867	1.042
Other meat preparations	0.796	0.916	0.415	0.389	0.390	0.428	0.434	0.435	0.440	0.588
Fish and seafood	1.594	1.727	0.444	0.368	0.371	0.379	0.380	0.382	0.384	0.545
Fresh or chilled fish	6.025	7.342	1.037	0.997	1.029	1.047	1.031	1.035	1.103	1.616
Frozen fish	2.067	2.462	1.310	0.358	0.396	0.341	0.310	0.310	0.320	0.414
Fresh or chilled seafood	2.523	2.836	1.570	1.538	1.543	1.568	1.551	1.557	1.783	2.551
Frozen seafood	1.958	2.720	2.060	0.537	0.524	0.510	0.509	0.509	0.570	0.769
Dried, smoked or salted fish and seafood	1.213	1.411	0.528	0.529	0.534	0.577	0.565	0.565	0.582	0.629
Other preserved or processed fish and seafood preparations	0.969	0.993	0.406	0.385	0.383	0.444	0.440	0.440	0.470	0.527
Milk, cheese and eggs	0.599	0.627	0.423	0.401	0.401	0.812	0.707	0.707	0.704	0.821
Whole and low fat milk	0.863	1.102	0.546	0.519	0.521	0.516	0.539	0.539	0.545	0.603
Preserved milk	0.757	0.870	0.648	0.639	0.638	0.788	0.778	0.778	0.805	0.977
Yoghurt	0.928	1.098	0.680	0.610	0.609	0.618	0.598	0.598	0.606	0.776
Ripening and cream cheese	0.803	0.772	0.550	0.511	0.512	0.520	0.551	0.552	0.582	0.578
Curd	0.832	1.026	0.378	0.383	0.382	0.424	0.425	0.426	0.423	0.573
Sour cream	0.784	0.880	0.577	0.518	0.516	0.577	0.559	0.560	0.588	0.679
Milk beverages and other dairy products	0.841	1.049	0.465	0.465	0.468	0.483	0.496	0.501	0.444	0.715
Eggs	2.358	2.023	2.162	1.988	1.985	5.248	4.262	4.264	4.292	5.034

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Oils and fats	0.923	1.113	0.926	0.912	0.913	1.048	1.069	1.072	1.094	1.319
Butter	1.472	1.892	1.629	1.575	1.574	1.659	1.676	1.676	1.737	1.798
Margarine and other vegetable fats	1.227	1.303	1.077	1.090	1.094	1.349	1.377	1.388	1.368	2.220
Olive oil	0.923	1.026	0.575	0.529	0.530	0.542	0.524	0.524	0.524	0.772
Other edible oils	1.043	1.062	0.711	0.658	0.658	0.786	0.729	0.731	0.757	0.944
Other edible animal fats	1.679	2.094	0.915	0.956	0.962	0.930	0.902	0.904	1.025	0.989
Fruits	2.121	2.158	1.960	1.886	2.249	2.318	2.110	2.116	2.610	3.316
Bananas	4.022	4.852	3.597	3.228	3.220	4.478	4.042	4.043	4.245	5.704
Apples	4.798	6.098	4.233	4.281	4.340	4.484	3.882	3.916	5.099	5.821
Berries	12.389	12.182	7.057	7.057	8.472	9.063	9.329	9.614	9.538	15.348
Fruits with a stone	11.848	12.028	5.319	5.538	5.856	6.555	7.140	7.242	8.293	10.765
Other fruits	6.159	6.611	3.116	3.579	3.833	4.747	5.252	5.315	6.712	8.135
Frozen fruit	1.847	2.008	0.946	0.863	0.861	0.888	0.878	0.884	0.898	1.367
Preserved fruit and fruit-based products	0.785	0.868	0.585	0.496	0.494	0.581	0.564	0.570	0.552	0.881
Vegetables	2.679	2.740	2.865	2.749	2.783	4.038	3.519	3.515	3.814	5.487
Lettuce	4.132	5.897	5.990	5.788	5.786	8.385	7.450	7.441	9.243	10.288
Cabbage	6.733	7.312	7.333	6.358	6.340	10.413	8.954	8.979	9.614	13.843
Cauliflower	6.590	7.198	7.190	6.731	6.773	11.757	10.921	11.121	9.938	19.357
Tomatoes	10.763	11.748	9.395	9.479	9.529	12.958	12.261	12.842	11.412	23.393
Cucumbers	14.302	14.917	16.227	15.616	15.759	22.515	21.122	21.257	20.620	31.924
Carrot	6.624	6.681	5.639	5.322	5.364	7.563	6.653	6.616	7.665	9.108
Beetroot	8.108	10.505	6.336	7.246	7.444	4.987	5.423	5.839	5.007	13.274
Onion	4.291	4.787	4.182	3.912	3.924	3.528	3.350	3.380	3.184	5.677
Other vegetables	3.409	3.403	2.844	2.766	2.779	3.613	3.187	3.183	3.657	5.595
Frozen vegetables other than potatoes and other tubers	1.019	1.324	0.481	0.475	0.476	0.492	0.484	0.486	0.518	0.620
Sauerkraut	2.607	2.752	1.989	1.767	1.765	2.521	1.818	1.860	2.955	2.469
Other tubers and products of tuber vegetables	1.153	1.158	0.789	0.720	0.737	0.818	0.876	0.906	0.895	1.751
Potatoes	16.966	16.805	15.477	13.659	13.713	15.637	11.440	11.664	12.487	26.223
Potatoes products	0.858	1.056	0.453	0.451	0.453	0.459	0.457	0.459	0.452	0.656
Crisps	0.973	1.115	0.521	0.553	0.553	0.537	0.558	0.563	0.550	0.814
Sugar, jam, honey, chocolate and others	0.532	0.660	0.493	0.494	0.498	0.595	0.652	0.662	0.630	0.780
Sugar	1.965	2.276	2.073	2.020	2.035	2.377	2.675	2.703	2.749	2.278
Jams, marmalades and honey	0.651	0.786	0.575	0.571	0.572	0.619	0.588	0.590	0.577	0.858
Confectionery products	0.588	0.658	0.330	0.340	0.342	0.387	0.420	0.427	0.421	0.702
Edible ices and ice cream	1.632	1.794	1.076	1.024	1.023	1.139	1.086	1.089	1.221	1.588
Artificial sugar substitutes	1.080	1.126	0.444	0.404	0.406	0.416	0.386	0.386	0.442	0.614
Sauces, condiments	0.509	0.563	0.351	0.359	0.360	0.335	0.342	0.343	0.334	0.531
Salt	1.155	1.446	0.618	0.766	0.783	0.669	0.630	0.630	0.648	0.790
Spices and culinary herbs	0.837	1.013	0.629	0.625	0.628	0.726	0.745	0.747	0.731	1.212
Baby food	0.638	0.763	0.410	0.400	0.401	0.482	0.447	0.447	0.547	0.545
Other food products n.e.c.	0.804	0.934	0.481	0.482	0.483	0.484	0.512	0.517	0.492	0.961
Non-alcoholic beverages	0.503	0.521	0.270	0.239	0.238	0.265	0.270	0.273	0.249	0.422
Coffee	0.880	0.925	0.427	0.400	0.400	0.448	0.415	0.415	0.415	0.548
Cocoa and powdered chocolate	1.013	1.249	0.702	0.671	0.671	0.758	0.771	0.775	0.764	1.105
Mineral or spring waters	0.844	0.916	0.503	0.453	0.453	0.463	0.476	0.478	0.461	0.764
Soft drinks	0.754	0.824	0.471	0.423	0.420	0.512	0.497	0.499	0.507	0.725
Fruit juices	0.789	0.838	0.762	0.740	0.740	0.883	0.846	0.849	0.865	1.251
Fruit and vegetables juices	0.907	1.024	0.549	0.527	0.527	0.547	0.547	0.550	0.564	0.807

Tab. 6 (Appendix) Relative RMSFE of eCPI and benchmarks nowcasts, Jan 2014 - Jun 2018.

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights
Food and non-alcoholic beverages	0.84 *	0.89	0.79 ***	0.76 ***	0.82 **
Bread and cereals	2.84 **	3.13 **	1.34 ***	1.84 *	1.92 **
Rice	0.84 *	1.00	0.75 ***	0.69 ***	0.68 ***
Wheat flours	0.80 ***	0.74 **	0.78 ***	0.73 ***	0.74 ***
Other flours	1.41 **	1.45 **	1.02	0.95	0.95
Groats and grains	1.68 **	2.10 ***	1.08	0.97	0.97
Bread	3.24 **	3.18 **	1.57 ***	2.21 **	2.33 **
Other bakery products	2.38 ***	2.66 ***	1.16 *	1.00	1.01
Pizza and quiche	1.99 ***	2.66 ***	0.92	0.94	0.94
Pasta products and couscous	1.48 **	2.00 **	0.97	0.94 *	0.95
Breakfast cereals	1.38 **	1.39 *	0.87 *	0.92	0.92
Other cereal products	1.62 ***	1.90 ***	0.71 ***	0.85 *	0.85 *
Meat	1.34 ***	1.55 ***	1.09	1.04	1.05
Beef	3.42 **	3.62 **	0.89	0.77 **	0.78 **
Veal	6.46 ***	7.85 ***	0.90	0.88 *	0.93
Pork	1.44 **	1.52 **	1.09	1.01	1.01
Lamb and goat	2.85 ***	3.35 ***	1.63 **	0.84	0.87
Chickens	1.02	1.01	0.87	0.89	0.89
Other poultry	3.81 *	3.82 *	0.93	0.84 ***	0.84 ***
Other meats	1.55 **	1.95 **	1.33	0.83 **	0.82 ***
Edible offal	3.79 **	4.67 **	1.15 **	1.04	1.06
Dried, salted or smoked meat other than poultry	2.44 ***	3.03 ***	1.20 ***	1.08	1.10
Dried, salted or smoked poultry meat	2.98 ***	3.85 ***	0.84 ***	0.78 ***	0.79 ***
Mixed ground meat	3.70 **	4.98 ***	0.96	0.91 **	0.93 **
Other meat preparations	1.86 ***	2.14 ***	0.97	0.91	0.91
Fish and seafood	4.20 **	4.55 **	1.17	0.97	0.98
Fresh or chilled fish	5.75 **	7.01 **	0.99	0.95	0.98
Frozen fish	6.07 ***	7.23 ***	3.85 **	1.05	1.16 **
Fresh or chilled seafood	1.61 ***	1.81 ***	1.00	0.98	0.98
Frozen seafood	3.84 ***	5.33 ***	4.04 *	1.05	1.03
seafood	2.10 ***	2.45 ***	0.91 *	0.92 *	0.93
Other preserved or processed fish and seafood preparations	2.18 ***	2.23 ***	0.91	0.87 **	0.86 **
Milk, cheese and eggs	0.74	0.77	0.52 *	0.49 *	0.49 *
Milk	1.68 **	2.15 ***	1.06	1.00	1.01
Whole and low fat milk	1.67 **	2.13 ***	1.06	1.01	1.01
Preserved milk	0.96	1.11	0.82 ***	0.81 ***	0.81 ***
Yoghurt	1.50 **	1.78 **	1.10 *	0.99	0.99
Ripening and cream cheese	1.54 ***	1.48 ***	1.06	0.98	0.98
Curd	1.96 ***	2.42 ***	0.89	0.90	0.90
Sour cream	1.36 **	1.52 ***	1.00	0.90	0.89
Milk beverages and other dairy products	1.74 *	2.17 *	0.96	0.96	0.97
Eggs	0.45 *	0.39 *	0.41 *	0.38 *	0.38 *

Notes: Asterisks indicate statistical significance of one-sided Diebold-Mariano test; * = p<0.1, ** = p<0.05, *** = p<0.01.

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights
Oils and fats	0.88	1.06	0.88 **	0.87 **	0.87 **
Butter	0.89	1.14	0.98	0.95	0.95
Margarine and other vegetable fats	0.91 *	0.97	0.80 **	0.81 ***	0.81 ***
Olive oil	1.70 ***	1.90 ***	1.06	0.98	0.98
Other edible oils	1.33 **	1.35 ***	0.90 **	0.84 ***	0.84 ***
Other edible animal fats	1.81 **	2.25 *	0.98	1.03	1.03
Fruits	0.91	0.93	0.85 **	0.81 **	0.97
Citrus fruits	0.78 **	0.92	0.79 ***	0.72 ***	0.72 ***
Bananas	0.90	1.08	0.80 **	0.72 **	0.72 **
Apples	1.07	1.36 *	0.94	0.95	0.97
Berries	1.37 **	1.34 **	0.78 **	0.78 **	0.93
Fruits with a stone	1.81 ***	1.83 ***	0.81 *	0.84 **	0.89 *
Other fruits	1.30 *	1.39 **	0.66 ***	0.75 ***	0.81 **
Frozen fruit	2.08 ***	2.26 ***	1.06	0.97	0.97
Dried fruit and nuts	1.26 *	1.45 **	1.05	0.93	0.92
Preserved fruit and fruit-based products	1.35 **	1.49 ***	1.01	0.85	0.85
Vegetables	0.66 ***	0.68 ***	0.71 ***	0.68 ***	0.69 ***
Lettuce	0.49 ***	0.70 **	0.71 ***	0.69 ***	0.69 **
Cabbage	0.65 **	0.70 **	0.70 *	0.61 **	0.61 **
Cauliflower	0.56 ***	0.61 ***	0.61 ***	0.57 ***	0.58 ***
Tomatoes	0.83 *	0.91	0.73 ***	0.73 ***	0.74 ***
Cucumbers	0.64 ***	0.66 **	0.72 **	0.69 **	0.70 **
Carrot	0.88	0.88	0.75 **	0.70 ***	0.71 ***
Beetroot	1.63 **	2.11 ***	1.27 *	1.45 **	1.49 **
Onion	1.22 *	1.36 **	1.19	1.11	1.11
Other vegetables	0.94	0.94	0.79 **	0.77 **	0.77 **
Frozen vegetables other than potatoes and other tubers	2.07 ***	2.69 ***	0.98	0.97 *	0.97
Sauerkraut	1.03	1.09	0.79 **	0.70 ***	0.70 ***
Other tubers and products of tuber vegetables	1.41	1.42	0.96	0.88 **	0.90 *
Potatoes	1.09	1.07	0.99	0.87 *	0.88 *
Potatoes products	1.87 ***	2.30 ***	0.99	0.98	0.99
Crisps	1.81 ***	2.07 ***	0.97	1.03	1.03
Sugar, jam, honey, chocolate and others	0.89	1.11	0.83 *	0.83 *	0.84 *
Sugar	0.83	0.96	0.87 *	0.85 *	0.86 *
Jams, marmalades and honey	1.05	1.27	0.93	0.92 **	0.92 **
Chocolate	1.15 *	1.50 ***	0.85	0.79 ***	0.79 **
Confectionery products	1.52 ***	1.70 ***	0.85 *	0.88 **	0.88 *
Edible ices and ice cream	1.43 **	1.57 ***	0.94	0.90 **	0.90 **
Artificial sugar substitutes	2.60 ***	2.71 ***	1.07 *	0.97	0.98
Food products n.e.c.	1.57 *	1.94 **	0.88 **	0.91 *	0.91 *
Sauces, condiments	1.52 ***	1.68 ***	1.05	1.07	1.07
Salt	1.73 ***	2.16 ***	0.92	1.14 **	1.17 **
Spices and culinary herbs	1.15	1.39	0.87	0.86 *	0.86 *
Baby food	1.33 **	1.58 ***	0.85 **	0.83 **	0.83 **
Ready-made meals	2.83 ***	3.48 ***	1.06	1.07	1.08
Other food products n.e.c.	1.66 ***	1.93 ***	0.99	0.99	1.00
Non-alcoholic beverages	1.90 ***	1.96 ***	1.02	0.90 *	0.90 *
Coffee	1.96 ***	2.06 ***	0.95	0.89 *	0.89 *
Tea	1.86 ***	2.00 ***	0.99	0.98	0.97
Cocoa and powdered chocolate	1.34	1.65 *	0.93	0.89 *	0.89 *
Mineral or spring waters	1.82 ***	1.98 ***	1.09	0.98	0.98
Soft drinks	1.47 ***	1.61 ***	0.92 *	0.83 ***	0.82 ***
Fruit juices	0.89	0.95	0.86 **	0.84 ***	0.84 ***
Fruit and vegetables juices	1.66 **	1.87 ***	1.00	0.96	0.96

Tab. 7 (Appendix) MFE of eCPI and benchmarks nowcasts, Jan 2014 - Jun 2018.

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Food and non-alcoholic beverages	0.021	0.033	0.149	0.156	0.201	0.258	0.299	0.299	0.318	0.017
Bread and cereals	-0.054	-0.024	0.009	0.053	0.055	0.054	0.106	0.108	0.116	-0.002
Rice	0.011	0.009	-0.051	-0.009	-0.008	0.077	0.100	0.099	0.059	-0.002
Wheat flours	-0.032	-0.059	0.055	0.079	0.080	0.243	0.218	0.220	0.258	-0.008
Other flours	0.089	0.048	0.235	0.243	0.246	0.122	0.176	0.182	0.118	0.009
Groats and grains	0.054	0.039	0.072	0.095	0.098	0.038	0.056	0.057	0.068	-0.012
Bread	-0.132	-0.058	-0.007	0.057	0.060	0.030	0.144	0.148	0.174	-0.005
Other bakery products	0.008	-0.003	-0.006	0.023	0.024	0.064	0.052	0.052	0.039	-0.002
Pizza and quiche	-0.047	-0.050	0.043	0.084	0.084	0.034	0.050	0.050	0.055	-0.002
Pasta products and couscous	0.037	0.070	0.107	0.102	0.102	0.090	0.084	0.084	0.063	0.021
Breakfast cereals	-0.017	-0.015	0.048	0.063	0.063	0.067	0.086	0.086	0.073	0.001
Other cereal products	0.066	0.095	0.044	0.059	0.063	0.113	0.080	0.080	0.077	-0.006
Meat	0.082	0.084	0.092	0.095	0.098	0.095	0.107	0.106	0.144	-0.017
Beef	0.104	0.075	0.018	0.087	0.096	0.122	0.164	0.164	0.151	0.000
Veal	0.230	0.329	0.048	0.091	0.099	0.111	0.170	0.171	0.198	-0.004
Pork	0.016	-0.007	0.137	0.099	0.098	0.186	0.131	0.130	0.190	-0.014
Lamb and goat	0.615	0.645	-0.015	0.071	0.079	0.241	0.228	0.226	0.235	0.001
Chickens	0.093	0.134	0.218	0.235	0.244	0.197	0.217	0.212	0.208	-0.073
Other poultry	0.013	0.072	0.149	0.137	0.145	0.131	0.126	0.125	0.100	-0.018
Other meats	0.295	0.349	0.459	0.161	0.149	0.097	0.082	0.081	0.096	-0.007
Edible offal	0.439	0.341	0.100	0.106	0.122	0.020	0.037	0.036	0.039	0.009
Dried, salted or smoked meat other than poultry	0.116	0.122	0.026	0.040	0.043	0.016	0.058	0.059	0.116	-0.008
Dried, salted or smoked poultry meat	-0.133	-0.129	0.048	0.049	0.047	0.048	0.038	0.037	0.050	-0.011
Mixed ground meat	-0.122	-0.086	0.069	0.062	0.060	0.115	0.105	0.104	0.084	-0.012
Other meat preparations	-0.027	-0.007	0.044	0.052	0.050	0.047	0.070	0.070	0.065	0.012
Fish and seafood	0.137	0.109	0.044	0.034	0.040	0.033	0.037	0.037	0.040	-0.010
Fresh or chilled fish	0.234	0.225	-0.030	-0.001	0.001	-0.030	0.028	0.028	0.122	-0.028
Frozen fish	-0.119	-0.079	0.147	0.133	0.129	0.064	0.027	0.026	-0.007	-0.009
Fresh or chilled seafood	0.507	0.685	0.306	0.232	0.226	0.285	0.235	0.232	0.155	0.011
Frozen seafood	-0.228	-0.096	0.058	0.048	0.041	0.044	0.039	0.038	0.063	0.013
Dried, smoked or salted fish and seafood	0.434	0.425	0.071	0.133	0.150	0.051	0.056	0.056	0.039	-0.007
Other preserved or processed fish and seafood preparations	0.016	0.018	0.006	0.031	0.035	0.041	0.036	0.036	0.011	-0.005
Milk, cheese and eggs	0.034	0.006	0.043	0.042	0.042	0.053	0.038	0.040	0.026	0.014
Milk	0.021	-0.008	0.074	0.067	0.068	0.094	0.108	0.108	0.082	0.010
Whole and low fat milk	0.023	-0.006	0.078	0.069	0.070	0.095	0.111	0.111	0.083	0.011
Preserved milk	-0.014	-0.047	0.002	0.027	0.028	0.073	0.061	0.060	0.050	0.000
Yoghurt	0.018	-0.028	0.011	0.012	0.009	-0.031	-0.053	-0.052	-0.038	0.015
Ripening and cream cheese	0.122	0.097	0.100	0.099	0.102	0.035	0.069	0.070	0.073	0.021
Curd	0.155	0.107	0.016	0.052	0.053	0.049	0.065	0.065	0.053	0.012
Sour cream	0.028	0.015	-0.046	0.027	0.029	-0.008	0.030	0.030	0.032	0.008
Milk beverages and other dairy products	-0.052	-0.116	0.007	0.011	0.009	0.028	0.021	0.020	0.043	0.005
Eggs	-0.161	-0.143	0.018	-0.054	-0.057	0.125	-0.078	-0.065	-0.155	0.018

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Oils and fats	0.116	0.126	-0.014	-0.006	-0.005	-0.025	-0.019	-0.019	0.061	-0.045
Butter	0.199	0.238	-0.031	-0.011	-0.011	-0.064	-0.049	-0.047	0.160	-0.062
Margarine and other vegetable fats	0.130	0.117	0.040	0.029	0.029	0.025	0.088	0.087	0.054	-0.019
Olive oil	-0.055	-0.067	-0.093	-0.094	-0.093	-0.030	-0.013	-0.012	-0.028	0.002
Other edible oils	0.083	0.063	0.081	0.081	0.083	0.097	0.034	0.034	0.027	-0.016
Other edible animal fats	0.131	0.092	0.042	0.061	0.074	0.066	0.110	0.112	0.204	-0.005
Fruits	-0.172	-0.138	0.316	0.328	1.064	0.355	0.372	0.385	0.537	0.055
Citrus fruits	0.140	0.226	0.061	0.101	0.101	-0.542	-0.502	-0.502	-0.685	-0.348
Bananas	0.138	0.262	0.595	0.423	0.414	0.809	0.901	0.897	1.079	0.227
Apples	0.526	0.443	0.120	0.222	0.252	0.301	0.526	0.546	1.464	0.002
Berries	-2.261	-2.193	-0.533	-0.387	4.151	0.032	-0.015	0.006	-0.341	0.848
Fruits with a stone	-2.102	-2.085	-0.163	-0.143	-0.181	-0.280	-0.572	-0.511	0.226	-0.435
Other fruits	-0.478	-0.579	-0.322	-0.570	-0.577	-0.029	0.160	0.164	0.475	0.127
Frozen fruit	0.178	0.246	0.165	0.201	0.202	0.120	0.087	0.087	0.011	0.017
Dried fruit and nuts	0.048	0.050	0.133	0.151	0.151	0.113	0.076	0.075	0.068	-0.004
Preserved fruit and fruit-based products	-0.018	-0.034	0.087	0.070	0.070	0.055	0.028	0.027	0.029	-0.011
Vegetables	-0.128	-0.034	0.803	0.774	0.798	1.798	1.980	1.969	1.981	0.219
Lettuce	-0.102	0.147	-0.054	-0.081	-0.078	3.180	2.142	2.092	2.317	0.632
Cabbage	-0.491	-0.364	0.358	0.471	0.483	0.655	1.418	1.415	2.760	0.076
Cauliflower	-0.621	-0.229	-0.304	-0.441	-0.443	-0.437	-0.279	-0.223	-0.789	1.189
Tomatoes	-1.412	-1.227	0.072	-0.101	-0.122	1.554	1.976	1.993	0.860	1.193
Cucumbers	-1.698	-0.686	-0.914	0.954	1.389	3.237	4.441	4.385	4.263	1.179
Carrot	-1.283	-1.273	-0.313	0.035	0.030	-1.507	-1.441	-1.433	-1.080	-0.548
Beetroot	-0.720	-0.867	-0.787	-0.761	-0.786	1.495	1.044	1.053	1.220	-0.412
Onion	-0.411	-0.315	0.126	-0.039	-0.044	0.260	0.159	0.146	0.122	-0.175
Other vegetables	-0.303	-0.312	0.007	-0.120	-0.120	0.566	0.544	0.533	0.954	0.062
Frozen vegetables other than potatoes and other tubers	0.153	0.214	0.130	0.110	0.111	0.112	0.091	0.090	0.085	0.014
Sauerkraut	-0.270	-0.253	0.238	0.125	0.119	0.683	0.544	0.539	0.959	-0.151
Other tubers and products of tuber vegetables	-0.015	-0.021	0.138	0.101	0.100	0.141	0.121	0.117	0.080	0.020
Potatoes	-2.483	-2.601	1.361	0.577	0.561	3.520	3.944	3.874	4.489	-0.505
Potatoes products	-0.052	-0.100	0.059	0.049	0.048	0.083	0.078	0.077	0.065	0.018
Crisps	-0.084	-0.059	0.072	0.109	0.109	0.101	0.101	0.099	0.115	0.008
Sugar, jam, honey, chocolate and others	-0.009	0.019	0.028	0.037	0.040	-0.001	0.197	0.204	0.174	-0.024
Sugar	0.073	0.163	0.310	0.282	0.289	0.121	0.991	1.028	0.906	-0.027
Jams, marmalades and honey	0.033	0.067	-0.037	0.068	0.071	0.055	0.046	0.046	0.052	-0.001
Chocolate	0.060	0.101	0.026	0.023	0.027	0.090	0.049	0.047	0.030	-0.036
Confectionery products	0.016	0.032	0.051	0.041	0.042	0.045	0.055	0.054	0.054	-0.018
Edible ices and ice cream	0.043	-0.001	0.093	0.103	0.103	0.035	0.053	0.053	0.029	0.016
Artificial sugar substitutes	-0.125	-0.153	-0.002	0.020	0.018	0.015	0.017	0.017	0.024	-0.011
Food products n.e.c.	-0.013	-0.015	0.086	0.083	0.084	0.071	0.088	0.088	0.054	-0.001
Sauces, condiments	0.086	0.069	0.095	0.098	0.098	0.057	0.072	0.072	0.048	-0.011
Salt	-0.034	0.001	0.117	0.178	0.182	0.016	0.058	0.057	0.040	0.000
Spices and culinary herbs	-0.017	-0.004	0.024	0.027	0.028	0.129	0.087	0.085	0.078	0.011
Baby food	-0.091	-0.081	0.139	0.128	0.127	0.074	0.109	0.108	0.024	0.002
Ready-made meals	-0.035	-0.006	0.069	0.092	0.091	0.059	0.073	0.072	0.057	0.018
Other food products n.e.c.	-0.044	-0.056	0.082	0.058	0.057	0.062	0.093	0.093	0.066	-0.019
Non-alcoholic beverages	0.066	0.080	0.036	0.049	0.050	0.056	0.039	0.039	0.032	-0.009
Coffee	0.003	0.038	-0.025	-0.008	-0.005	0.018	-0.003	-0.003	-0.002	-0.014
Tea	-0.003	-0.002	0.058	0.072	0.072	0.032	0.040	0.040	0.026	-0.012
Cocoa and powdered chocolate	-0.018	-0.016	0.132	0.139	0.138	0.084	0.058	0.058	0.085	-0.017
Mineral or spring waters	0.038	0.030	0.017	0.018	0.019	0.070	0.050	0.049	0.043	-0.014
Soft drinks	0.183	0.219	0.115	0.125	0.125	0.114	0.081	0.080	0.069	0.000
Fruit juices	0.056	0.023	-0.022	0.021	0.021	0.005	0.017	0.017	0.004	-0.013
Fruit and vegetables juices	0.063	0.054	0.097	0.085	0.085	0.039	0.069	0.069	0.045	0.018

Tab. 8 (Appendix) MAFE of eCPI and benchmarks nowcasts, Jan 2014 - Jun 2018.

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Food and non-alcoholic beverages	0.355	0.371	0.329	0.320	0.342	0.443	0.423	0.421	0.474	0.570
Bread and cereals	0.324	0.359	0.190	0.204	0.212	0.142	0.181	0.188	0.174	0.178
Rice	0.664	0.768	0.616	0.566	0.562	0.856	0.837	0.838	0.860	0.991
Wheat flours	0.905	0.772	0.862	0.811	0.823	1.097	1.052	1.054	1.201	1.491
Other flours	1.027	1.049	0.781	0.709	0.707	0.733	0.737	0.737	0.766	0.965
Groats and grains	0.498	0.594	0.361	0.321	0.319	0.339	0.345	0.345	0.373	0.384
Bread	0.496	0.525	0.318	0.356	0.371	0.179	0.245	0.259	0.246	0.190
Other bakery products	0.443	0.503	0.245	0.210	0.212	0.212	0.219	0.220	0.225	0.292
Pizza and quiche	0.599	0.847	0.288	0.280	0.279	0.317	0.299	0.301	0.305	0.493
Pasta products and couscous	0.565	0.675	0.399	0.384	0.385	0.413	0.419	0.421	0.438	0.558
Breakfast cereals	0.365	0.352	0.220	0.238	0.238	0.265	0.277	0.277	0.284	0.344
Other cereal products	0.797	0.873	0.350	0.419	0.421	0.485	0.460	0.460	0.467	0.544
Meat	0.456	0.507	0.345	0.326	0.328	0.350	0.303	0.306	0.360	0.497
Beef	1.054	1.199	0.348	0.301	0.307	0.399	0.406	0.409	0.396	0.356
Veal	1.197	1.378	0.222	0.218	0.227	0.241	0.283	0.287	0.273	0.243
Pork	1.089	1.154	0.873	0.813	0.812	0.829	0.752	0.752	0.853	1.010
Lamb and goat	1.832	2.073	0.899	0.575	0.594	0.630	0.635	0.634	0.629	0.858
Chickens	1.584	1.535	1.273	1.331	1.340	1.543	1.295	1.306	1.439	2.294
Other poultry	1.544	1.649	0.674	0.616	0.626	0.757	0.679	0.681	0.758	0.832
Other meats	1.712	1.930	1.356	1.029	1.018	1.230	1.173	1.175	1.318	1.546
Edible offal	0.969	1.142	0.372	0.346	0.355	0.314	0.310	0.310	0.479	0.350
Dried, salted or smoked meat other than poultry	0.576	0.680	0.290	0.260	0.261	0.241	0.218	0.217	0.270	0.284
Dried, salted or smoked poultry meat	0.965	1.140	0.296	0.274	0.279	0.345	0.298	0.299	0.295	0.367
Mixed ground meat	2.161	2.732	0.621	0.620	0.630	0.680	0.645	0.647	0.699	0.846
Other meat preparations	0.648	0.734	0.341	0.325	0.326	0.350	0.350	0.351	0.363	0.474
Fish and seafood	0.878	0.959	0.316	0.290	0.289	0.295	0.295	0.296	0.300	0.400
Fresh or chilled fish	2.814	3.288	0.800	0.750	0.778	0.786	0.762	0.765	0.866	1.166
Frozen fish	1.425	1.728	0.696	0.277	0.304	0.264	0.239	0.238	0.268	0.321
Fresh or chilled seafood	1.882	2.004	1.136	1.112	1.121	1.146	1.138	1.135	1.381	1.876
Frozen seafood	1.389	1.691	0.990	0.413	0.409	0.411	0.408	0.407	0.452	0.610
Dried, smoked or salted fish and seafood	0.966	1.046	0.432	0.424	0.430	0.466	0.457	0.457	0.470	0.505
Other preserved or processed fish and seafood preparations	0.761	0.769	0.306	0.305	0.305	0.328	0.330	0.330	0.351	0.391
Milk, cheese and eggs	0.426	0.464	0.301	0.278	0.278	0.400	0.369	0.369	0.373	0.419
Whole and low fat milk	0.617	0.774	0.430	0.417	0.418	0.420	0.419	0.418	0.432	0.472
Preserved milk	0.585	0.662	0.444	0.432	0.432	0.577	0.572	0.573	0.579	0.664
Yoghurt	0.724	0.802	0.556	0.499	0.497	0.515	0.503	0.503	0.509	0.667
Ripening and cream cheese	0.631	0.598	0.447	0.405	0.405	0.448	0.453	0.454	0.484	0.493
Curd	0.692	0.869	0.299	0.307	0.307	0.325	0.335	0.337	0.329	0.455
Sour cream	0.603	0.696	0.443	0.421	0.419	0.457	0.444	0.445	0.462	0.536
Milk beverages and other dairy products	0.609	0.724	0.372	0.381	0.383	0.385	0.393	0.397	0.349	0.578
Eggs	1.090	1.078	1.257	1.116	1.109	1.892	1.587	1.601	1.852	1.842

Name	eCPI ex post	eCPI (real-time)	eCPI-in-ADL (best)	eCPI combination (1/RMSE)	eCPI combination equal weights	Benchmarks (best)	benchmarks combination (1/RMSE)	benchmarks combination equal weights	ARMA AIC	RW
Oils and fats	0.657	0.738	0.579	0.558	0.560	0.729	0.752	0.755	0.765	0.984
Butter	1.069	1.317	1.083	1.022	1.022	1.159	1.128	1.128	1.238	1.311
Margarine and other vegetable fats	0.868	0.952	0.816	0.813	0.813	1.038	1.033	1.040	1.018	1.613
Olive oil	0.708	0.782	0.458	0.429	0.430	0.435	0.421	0.419	0.424	0.613
Other edible oils	0.821	0.833	0.600	0.548	0.547	0.651	0.587	0.588	0.633	0.761
Other edible animal fats	1.185	1.290	0.737	0.677	0.672	0.758	0.694	0.693	0.822	0.750
Fruits	1.554	1.681	1.558	1.546	1.723	1.918	1.714	1.714	2.161	2.747
Citrus fruits	2.574	3.075	2.389	2.299	2.317	3.201	3.140	3.202	2.981	4.634
Bananas	3.148	3.472	2.877	2.535	2.530	3.493	3.304	3.300	3.513	4.593
Apples	3.621	4.423	3.231	3.209	3.246	3.322	3.048	3.065	3.928	4.070
Berries	9.209	9.414	5.649	5.609	6.915	6.801	7.364	7.594	7.588	12.168
Fruits with a stone	8.166	8.300	3.691	3.825	4.021	5.014	4.996	5.069	6.169	7.050
Other fruits	4.575	4.680	2.576	2.777	2.930	3.671	4.075	4.128	5.040	6.437
Frozen fruit	1.318	1.495	0.761	0.725	0.724	0.703	0.703	0.708	0.709	1.072
Dried fruit and nuts	0.695	0.751	0.537	0.493	0.493	0.529	0.504	0.506	0.496	0.814
Preserved fruit and fruit-based products	0.600	0.679	0.452	0.379	0.377	0.456	0.441	0.446	0.447	0.697
Vegetables	2.111	2.103	2.124	2.124	2.158	3.259	2.905	2.902	3.154	4.173
Lettuce	3.362	4.683	4.663	4.408	4.400	6.213	5.580	5.639	7.418	7.464
Cabbage	5.248	5.725	5.575	4.536	4.536	7.648	6.619	6.673	7.253	10.110
Cauliflower	4.956	5.800	5.582	5.245	5.283	9.646	8.958	9.039	8.356	15.761
Tomatoes	8.808	9.551	7.364	7.129	7.163	9.899	8.799	9.174	8.236	17.640
Cucumbers	10.102	10.853	12.167	11.456	11.694	16.846	15.539	15.610	15.009	24.887
Carrot	4.557	4.638	4.137	3.990	4.035	5.547	4.638	4.626	5.403	6.595
Beetroot	5.446	7.368	4.401	4.959	5.083	3.876	3.968	4.324	3.806	8.036
Onion	3.081	3.434	3.036	2.730	2.724	2.705	2.492	2.494	2.532	4.094
Other vegetables	2.620	2.618	2.159	2.157	2.168	2.749	2.379	2.390	2.948	4.336
Frozen vegetables other than potatoes and other tubers	0.762	0.961	0.394	0.376	0.376	0.389	0.372	0.372	0.393	0.484
Sauerkraut	1.924	2.137	1.486	1.290	1.290	1.999	1.290	1.288	2.079	1.963
Other tubers and products of tuber vegetables	0.866	0.849	0.558	0.512	0.531	0.607	0.617	0.638	0.639	1.259
Potatoes	8.990	8.853	7.627	7.626	7.724	9.006	7.278	7.700	8.353	12.820
Potatoes products	0.622	0.738	0.355	0.349	0.350	0.382	0.375	0.377	0.364	0.535
Crisps	0.777	0.870	0.427	0.454	0.454	0.432	0.459	0.463	0.436	0.663
Sugar, jam, honey, chocolate and others	0.436	0.533	0.419	0.412	0.415	0.450	0.494	0.502	0.491	0.614
Sugar	1.544	1.782	1.627	1.599	1.620	1.732	1.904	1.912	1.981	1.593
Jams, marmalades and honey	0.469	0.552	0.461	0.466	0.467	0.488	0.472	0.474	0.451	0.706
Chocolate	0.621	0.773	0.442	0.424	0.425	0.528	0.501	0.505	0.497	0.932
Confectionery products	0.449	0.511	0.255	0.267	0.270	0.309	0.342	0.347	0.338	0.578
Edible ices and ice cream	1.228	1.328	0.862	0.805	0.804	0.918	0.874	0.878	0.982	1.275
Artificial sugar substitutes	0.815	0.904	0.349	0.313	0.315	0.338	0.305	0.305	0.342	0.474
Food products n.e.c.	0.315	0.405	0.206	0.212	0.213	0.239	0.250	0.253	0.232	0.402
Sauces, condiments	0.399	0.456	0.285	0.294	0.295	0.273	0.273	0.275	0.271	0.434
Salt	0.860	1.045	0.516	0.601	0.613	0.536	0.497	0.497	0.520	0.649
Spices and culinary herbs	0.601	0.718	0.499	0.457	0.458	0.572	0.565	0.568	0.563	0.983
Baby food	0.500	0.571	0.317	0.290	0.290	0.359	0.338	0.338	0.426	0.406
Ready-made meals	0.713	0.869	0.289	0.299	0.301	0.277	0.277	0.278	0.288	0.507
Other food products n.e.c.	0.599	0.746	0.389	0.385	0.386	0.362	0.390	0.396	0.370	0.743
Non-alcoholic beverages	0.399	0.398	0.231	0.208	0.207	0.226	0.226	0.227	0.207	0.343
Coffee	0.662	0.714	0.352	0.335	0.337	0.379	0.337	0.337	0.328	0.447
Tea	0.574	0.658	0.300	0.292	0.292	0.316	0.309	0.312	0.338	0.483
Cocoa and powdered chocolate	0.673	0.841	0.534	0.483	0.484	0.552	0.561	0.563	0.566	0.853
Mineral or spring waters	0.655	0.726	0.402	0.369	0.369	0.381	0.389	0.389	0.370	0.594
Soft drinks	0.605	0.661	0.379	0.353	0.351	0.431	0.417	0.419	0.421	0.569
Fruit juices	0.617	0.632	0.618	0.555	0.554	0.668	0.632	0.636	0.649	0.920
Fruit and vegetables juices	0.663	0.773	0.444	0.435	0.435	0.441	0.429	0.431	0.448	0.608

www.nbp.pl