WHAT DRIVES THE DIFFERENCE BETWEEN ONLINE AND OFFICIAL PRICE INDEXES?

OLEKSANDR FARYNA[a], OLEKSANDR TALAVERA[b], TETIANA YUKHYMENKO[c] [*]

[a]National Bank of Ukraine, National University of "Kyiv-Mohyla Academy"
 Corresponding author
 Email: oleksandr.faryna@bank.gov.ua
[b]Swansea University, UK
 Email: o.talavera@swansea.ac.uk
[c]National Bank of Ukraine
 Email: tetiana.yukhymenko@bank.gov.ua

Abstract	This paper examines the associations between online price indexes and official statistics. First, we generate online CPI component sub-indexes, which are later aggregated to an Online Price CPI. This approach is applied to our unique dataset which contains about 3 million observations of online retail prices for consumer goods in Ukraine's five largest cities. The data span the period 2016m1 – 2017m12 and cover about 46% of Ukraine's Consumer Price Inflation basket. We find that online inflation is generally consistent with official estimates, but the matching capability varies across sub-indexes. Although the differences can partially be explained by poor dataset coverage, we find that online prices may indeed represent new information that is not captured by official statistics.
JEL Codes	C55, E31, E37
Keywords	online prices, web scraping, consumer price index, micro prices, big data

1. INTRODUCTION

One of the key aims of central banks is to keep inflation rates low. Inflation targeting has thus become a mainstream policy approach among policy makers in recent decades (e.g., Hammond, 2011; Jahan, 2017; Roger, 2010). However, in order to "target" inflation, central bankers need an explicit and observable measure of inflation that can serve as the nominal anchor for society. Measuring inflation is not always a straightforward exercise and it is usually beyond the remit of central banks. In particular, central banks commonly refer to publicly available indicators officially published by state statistics agencies, such as the consumer price index (CPI).

The CPI is the indicator usually selected by central banks, as it measures the cost of living in the economy and is easily accessible to both the public and policy makers. Despite its simplicity and public acceptance, it may not be the best measure: CPI covers a limited number of goods and services in the economy and might not capture overall inflation developments as perceived by the public. This may affect the effectiveness of the central bank's decisions and

question the success of monetary policy in general. Policy makers, therefore, should be armed with all possible tools and use all available sources of information to improve their ability to recognize and understand threats to price stability.

The rapid emergence of e-commerce in the retail sector has made it possible to observe the prices for various goods and services online. Web scraping (collecting data from online sources through the use of specially written software), has become a useful tool for gathering data on online prices from the web in order to complement official statistics. Many national statistics organizations and other public institutions have already launched web-scraping projects to improve their data collection process, including the U.S. Bureau of Labor Statistics (Horrigan, 2013), the U.K. Office of National Statistics (Breton et al., 2015), Statistics Netherlands (Griffioen, de Haan, Willenborg, 2014), Statistics New Zealand (Krsinich, 2015), and Statistics Norway (Nygaard, 2015). Compared to other methods of data collection, web scraping has a range of advantages: in addition to the low cost of data collection, scrapped data are available on real-time and high frequency

basis, which can help policy makers to continually monitor inflation developments at the micro-level.

A growing literature utilizes online data for research purposes. The Billion Price Project (BPP)[1] founded in 2008 at MIT by Alberto Cavallo and Roberto Rigobon, aims to collect prices from hundreds of online retailers around the world. Cavallo and Rigobon (2016) show that online prices can be successfully used as an alternative source of information for constructing consumer price indexes. Some studies use online data to check whether official statistics is accurate and that it has not been manipulated. In particular, Cavallo (2013) uses online prices to study how online indexes match up with official statistics in five Latin American countries. The author finds that while for Brazil, Chile, Columbia, and Venezuela online price indexes approximate both the level and the main dynamics of official inflation, Argentina's web inflation was nearly three times higher than official statistics. Coupe and Petrusha (2014), in turn, find that online and official consumer inflation in Ukraine may differ considerably in the short term, but the deviation can be both positive and negative.

Comparing online and official price indexes, it is important to understand why potential differences may occur. On the one hand, online prices may indeed represent new information about long-term inflation developments that are not captured by official statistics. Meanwhile, given that online markets tend to be more flexible, online prices can adjust to new economic conditions more quickly,[2] hence producing short-term deviations, while in the long-run online inflation should be consistent with official estimates. On the flip side, differences may arise due to technical issues, fundamentally different approaches in data collection, and the methods used to construct online indexes. In contrast to official data, web-scraped data usually includes a high number of goods items, while the coverage of retailers and regions is limited. In addition, the high frequency of sampling for the online dataset often results in a high number of missing observations, due to errors in the scraping scripts or simply because goods may be out of stock. As a result, the composition of goods included in online price indexes can vary dramatically over time, which is usually inconsistent with the standard approaches used by statistics organizations. So before coming to any conclusions about whether online prices reflect new information about inflation developments that are not captured by official statistics, it is important to explore what drives such differences.

In this paper we develop an online consumer price index for Ukraine using a rich dataset of online prices and compare it to the official statistics reported by the State Statistics Service of Ukraine. We generate online CPI component sub-indexes and, thereafter, aggregate them into an Online Price CPI. Our unique dataset contains about 3 million observations of online retail prices for consumer goods in the five largest cities in Ukraine. The data span the period 2016m1 – 2017m12 and cover about 46% of Ukraine's Consumer Price Inflation basket. We find that online inflation is in general consistent with official estimates, but the matching capability varies across sub-indexes. We further explore those properties of the dataset that can account

for the differences. For this purpose, we employ alternative filtering and aggregation techniques that improve or reduce the matching performance of the constructed indexes. We find that online price indexes may deviate from their official counterparts because of technical issues in data collection and poor dataset coverage. However, our analysis indicates that online prices can outpace reported estimates and convey new information that is not captured by the official CPI.

The rest of the paper proceeds as follows. The second section describes the online dataset used for our analysis. The third section introduces sub-component as well as aggregated consumer price online indexes and explores their ability to match official statistics. The fourth section gives conclusions.

2. ONLINE PRICES FOR CONSUMER GOODS IN UKRAINE

Our analysis utilizes online prices for consumer goods in Ukraine obtained from web scraping performed by the National Bank of Ukraine (NBU). In 2015, the NBU launched a web-scraping project aimed at improving data collection on consumer prices and at complementing official CPI statistics.

The Consumer Price Index provided by the State Statistics Service of Ukraine is the major indicator for tracking inflation developments used by the National Bank of Ukraine in the conduct of its monetary policy. The Ukrainian CPI basket comprises 328 sub-components, with up to 40% being food items, beverages and alcohol. Table 1 provides descriptive statistics for the Headline CPI and major CPI aggregates.

The NBU's online dataset includes several leading online retailers, which in addition to online stores have a wide network of offline supermarkets around the country in five major cities (Kyiv, Kharkiv, Dnipro, Odesa, and Lviv). These supermarkets and their online platforms offer a wide range of food items, beverages, alcohol, and tobacco products. The dataset covers up to 46% of the CPI basket and more than 130 CPI sub-components. Since the beginning of the project, the NBU's dataset has included over 75,000 goods items, with up to 3 million weekly observations[3] over two years (2016m1 – 2017m12). Most of the online prices are those of goods sold online in the Kyiv region, which could be considered the largest consumer in terms of e-commerce. Kharkiv, Dnipro, and Odesa have approximately equal shares, while Lviv is barely represented in the dataset so far. Table 2 provides descriptive statistics of the dataset.

3. CONSTRUCTING ONLINE INDEXES

The NBU's online dataset provides extended information on prices for goods at the micro level in various regions of Ukraine. In order to explore whether online price inflation is consistent with official statistics, we proceed by constructing online indexes and comparing them with their official counterparts.

---

[1] See for instance: http://www.thebillionpricesproject.com
[2] See, for example, Gorodnichenko & Talavera (2017).
[3] The data is collected daily, but we use weekly observations which are obtained by taking the mean price over the week. This helps us to avoid problems of an excessive number of missing observations and temporary errors in the web-scraping scripts.

3.1. Sub-Component Online Indexes

Following common practice,[4] we construct online sub-indexes as simple averages of week-on-week price changes within a narrowly defined group, namely at the CPI sub-component level:

$$\Delta p_{i,t} = \sum_{j=1}^{K} \left[\frac{(P_{ij,t} - P_{ij,t-1})}{P_{ij,t-1}} \right] \div K,$$

where Δp, $i=1,2,3...N$, states for an average week-on-week percentage change of prices at the sub-component level i; $P_{ij,t}$, $j=1,2,3...K$, is the price for a specific good j at the sub-component level i.

Thereafter, the sub-component weekly data series are transformed into a monthly frequency data set, so as to be comparable with official statistics. Since the dataset is of the weekly frequency and the number of weeks differs across each month, we first transform the web data in order to have four observations over the month, which prevents frequency conversion problems. This is done by dividing the month into four parts and matching the web data (e.g., 1st seven days, 2nd seven days, 3rd seven days, and the rest of the days). If there is more than one observation within a particular period in a month, they are simply averaged. Thereafter, we generate month-on-month indexes and convert weekly series to monthly ones:

$$\Delta_4 p_{i,m}^w = \prod_{j=1}^{4} (\Delta p_{i,m+w-j} + 1),$$

where $\Delta_4 p_{i,m}^w$ states for month-on-month change of online prices at week $w=[1:4]$. As a result, we get four monthly series representing month-on-month price changes.

Figure 1 plots several online sub-indexes together with their official counterparts. We present online indexes constructed at the third week of each month, since the State Statistic Service of Ukraine declares that it collects prices approximately at the same time. The matching performance of online prices varies across indexes. For example, online price indexes for eggs, apples, grapes, and kefir closely approximate both the trend dynamics and the short-term changes in the official statistics, with the errors not exceeding two standard deviations. Some indexes, e.g., for a loaf of bread, frozen fish, and sunflower oil, capture the trend dynamics of monthly inflation, but can differ considerably in the short-term. The deviation of online indexes from official data for beef tenderloin and chocolate, in turn, can be more sustained in some periods.

In addition to the visual inspection, we test the matching performance of the online indexes by calculating the Root Mean Square Error (RMSE) at the sub-component CPI level (see Table 3). For comparison purposes, we also provide RMSEs adjusted to a specific category's standard deviation of official inflation, since the volatility of sub-indexes can differ dramatically. We provide calculations for four weekly month-on-month indexes. The results suggest that even adjusted RMSEs can vary considerably across sub-indexes. The mean RMSEs exceed the official inflation standard

deviation by two times. While the minimum adjusted RMSE is around 0.5%, the maximum value is over 11%. Meanwhile, in about 70% of online sub-indexes, the overshoot errors dominate those that undershoot official estimates.

In order to determine what drives the differences between the online and official price indexes, we apply various filtering techniques and explore the properties of the data which improve or reduce the performance of the online indexes. For this purpose, we construct alternative online indexes by randomly excluding goods from the dataset. In particular, we run 99 iterations in which each good has a probability of 1%, 2%, and up to 99% of remaining in the dataset. For each probability level, we repeat the procedures 100 times and, consequently, we obtain 9,900 alternative datasets with different compositions of goods. For each alternative dataset, we construct four monthly sub-component online indexes as described above. We compare the generated indexes to official statistics by calculating their RMSEs. Given that each alternative dataset comprises different numbers of goods with a different number of missing observations and a unique mean standard deviation, we can now explore which of the features of the dataset affect the matching performance of the constructed online indexes. For this purpose, we estimate a panel regression of the following form:

$$RMSE_{i,k} = \beta_o + \beta_1 Obs_{i,k} + \beta_2 Members_{i,k} + \\ + \beta_3 SD_{i,k} + u_i + \varepsilon_{ik},$$

where $RMSE_{i,k}$ refers to an average RMSE over four weekly m-o-m sub-indexes i for iteration k; $Obs_{i,k}$, $Members_{i,k}$, and $SD_{i,k}$ state for the average number of non-missing observations, the number of sub-component group members, and the group's mean standard deviation, respectively; u_i refers to a cross-component time-invariant fixed effect, which allows the capture of component-specific performance; finally ε_{ik} is an error term.

The coefficient estimates obtained (see Table 4) suggest that the higher the volatility of online prices at the sub-component level, the lower the matching performance of the constructed indexes. Meanwhile, the lower the number of goods in the dataset, the higher the RMSE and, hence, the less accurate is the matching performance. This suggests that the differences between online and official inflation can be caused by a poor dataset. A less intuitive result is obtained for the average number of non-missing observations, indicating a positive correlation with the forecast error. Namely, the more observations in a sample, the higher is its RMSE.

We proceed by filtering-out goods that lower the performance of the web indexes. First, we focus on excluding goods that are characterized by high standard deviations within a narrowly defined group at the sub-component level. We construct alternative web indexes by excluding the upper and lower percentiles of standard deviations (e.g. 121 iterations starting from 0 to the 50th upper and lower percentile) and calculate the share of indexes with RMSEs below average over all iterations. Figure 2 provides the results of this exercise.

---

[4] Our approach is similar to Cavallo (2013) but we use a simple mean of price changes instead of a geometric mean.

Indeed, as suggested by the panel regression analysis presented above, excluding up to the 20th upper percentile of goods in a narrowly-defined group with high standard deviations improves the performance of web indexes. In contrast, the exclusion of the lower percentile does not seem to lower RMSE. On the one hand, this may indicate that the online dataset includes some outliers that, for technical reasons, are collected by the web-scraping procedure. In particular, a price can change dramatically due to changes in quantities. If a retailer changes the quantity of a good but uses the same web page, the web-scraping scripts are not able to recognize this without interfering. Nevertheless, one would expect that the share of such outliers caused by technical issues should be small. In our case, however, this share can exceed 20 percent, indicating that highly volatile prices may indeed represent new information in the short-term that is not captured by official statistics.

In order to check whether the number of non-missing observations influences the forecast error, we repeat the above procedure by excluding goods with a high number of missing observations. However, the results of the panel regression are confirmed, as the mean RMSE rises with the increase of the filtering requirements. Given that the number of group members has a negative correlation with the RMSE, additional exclusion of goods worsens the performance of the web indexes. These results suggest that including goods with a high number of observations over the sample does not necessarily guarantee better performance and, therefore, the inclusion of rarely tracked goods with a high number of missing observations does not reduce the matching performance.

Finally, we test the matching performance of online indexes generated at different weeks over the month. This exercise aims to explore the capability of online indexes to match official statistics as new data appears. In addition to the weekly indexes which incorporate price dynamics over the last four weeks (e.g. m-o-m price changes), we calculate average online inflation as the time passes. For instance, at the end of the first week of the month we have information on how prices changed compared to the first week of the previous month. At the end of the second week, in addition to first week inflation, we obtain data on the second week inflation. In order to better capture price dynamics, we can also calculate the average of the first and second weeks' inflation. The same applies to the subsequent weeks. We also compare official inflation for a particular month with the last week online inflation of the previous month and the first week of the preceding month. Figure 3 presents the resulting sample's mean RMSE for different weekly month-on-month indexes.

The results suggest that the share of month-on-month online indexes with the lowest RMSEs is the highest for second-week online inflation. It is noteworthy that the State Statistics Service of Ukraine collects price data at the beginning of the second half of the month. In addition, cross-component mean RMES for online indexes which comprise the average of the last month-on-month indexes decreases as the time passes and new web data arrives. This provides additional confirmation that online inflation may outpace official estimates and, therefore, have a predictive power.

To sum up, our analysis suggests that online inflation is in general consistent with official statistics, although the matching performance differs across CPI sub-indexes. The

differences can be explained both by the properties of the dataset, e.g., such as pure goods coverage, and by the fact that online prices indeed represent new information not captured by official statistics. In particular, online prices might be much more volatile and react more quickly to new economic conditions.

3.2. Aggregated Online Indexes

In the previous section we constructed online consumer price indexes at the sub-component level. Herein, we proceed by constructing the headline consumer price web index along with other CPI aggregates in order to explore how web data can approximate overall price developments in the country.

We employ several alternative approaches to constructing an aggregated online index. First, we use a simple average of all web indexes that represent a specific category. In particular, for aggregated headline inflation, we use a simple average of all online indexes, while for food online inflation we include only those indexes that belong to the food category. Alternatively, we use the officially available weights of the CPI basket structure provided by the State Statistics Service of Ukraine. Since the web dataset comprises up to 46% of the CPI basket (e.g. 134 out of 328 components) we construct relative weights using only those components that are represented in the dataset. Finally, to benchmark our results, we construct an index that comprises the average price dynamics of all goods in the dataset without constructing sub-component indexes. Table 5 reports the RMSEs of aggregated web indexes for headline CPI, constructed headline CPI (e.g., including only the components presented in the web dataset), food CPI and specific food indexes, beverages CPI, as well as alcohol and tobacco CPI. Figure 4 illustrates our results and plots official and web inflation.

For most aggregated web indexes, a weighted average of the sub-component indexes seems to improve the performance of the web data. In particular, for the constructed headline CPI, which includes only those components presented in the online dataset, the RMSE decreases from 1.06% to 0.93%. Similar applies to aggregated food indexes, as our web dataset covers most of the food sub-components.

Although, the results for the aggregated online indexes of headline CPI are mixed, the RMSE does not exceed 1%. This indicates that while the share of sub-components that are not presented in the dataset play an important role, our web dataset captures the overall price developments in the country, as the root mean square error for most indexes does not exceed one standard deviation of official statistics.

4. CONCLUSIONS

The rapid development of e-commerce over the last few decades has allowed policy-makers to enrich their toolbox for observing current developments in the economy using big data. In this paper, we construct an online consumer price index using a rich dataset of online prices obtained from webscraping performed by the National Bank of Ukraine and compare it to official statistics. We first generate sub-component online indexes and, thereafter, we aggregate them to the headline CPI index, as well as other CPI categories. Our dataset contains about 3 million

observations of online retail prices for consumer goods in the five largest cities in Ukraine and spans the period 2016m1 – 2017m12. The online data cover about 46% of Ukraine's CPI basket.

We explore which properties of the dataset improve or reduce its capability to approximate official statistics. Our findings suggest that online price indexes are in general consistent with official statistics, but the matching performance of online data varies across different CPI sub-components. The differences are partially explained by technical features of the dataset. In particular, the number of goods in the dataset matters, suggesting that the capability of online indexes to match official statistics increases when the online dataset covers a wide range of goods in a narrowly defined group. In contrast, goods with high number of observations over the sample do not necessarily guarantee a better fit, suggesting that the inclusion of rarely tracked goods with a high number of missing observations does not affect the matching performance of online indexes. Finally, utilizing officially provided CPI weights in constructing aggregated indexes decreases the deviation of online price indexes from their officially provided counterparts.

On the flip side, online indexes may indeed represent new information not captured by official statistics. The online prices of some goods may be much more volatile and, consequently, excluding such goods increases the matching performance of online indexes. The ability of high frequency online data to approximate official monthly inflation increases when a broader period of online price changes is taken into account. This indicates that online prices may react to new economic conditions more quickly and, consequently, have some predictive power for official statistics.

Our analysis confirms growing evidence in the literature (e.g., Cavallo and Rigobon, 2016; Breton et. al., 2015) that online prices can be used as an additional source of information for observing current developments in the inflation environment. This might also be relevant for so-called nowcasting or short-term forecasting, since online data is available in real-time and on a high-frequency basis. Therefore, our further research will consider the development of an inflation nowcasting framework that utilizes online data together with more traditional approaches on nowcasting. In particular, online prices might potentially improve the performance of dynamic factor models, which are commonly used to nowcast macroeconomic indicators.

REFERENCES

• Breton R., Clews G., Metcalfe L., Milliken N., Payne C., Winton J., Woods A. (2015). Research Indexes Using Web Scraped Data. Office for National Statistics, UK.

• Cavallo A., Rigobon R. (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. Journal of Economic Perspectives, Vol. 30, No. 2, pp. 151-178.

• Cavallo A. (2013). Online and Official Price Indexes: Measuring Argentina's Inflation. Journal of Monetary Economics, Vol. 60, Issue 2, pp. 152-165.

• Coupe T., Petrusha E. (2014). Can We Trust Official Inflation Measures? A Check Based on Inflation at Ukrainian Online Supermarkets. Focus Ukraine.

• Griffioen R., Haan J., Willenborg L. (2014). Collecting Clothing Data from the Internet. Proceedings of Meeting of the Group of Experts on Consumer Price Indexes, May 26–28.

• Gorodnichenko Y., Talavera O. (2017). Price Setting in Online Markets: Basic Facts, International Comparisons, and Cross-Border Integration. American Economic Review, Vol. 107, No. 1, pp. 249-282.

• Hammond G. (2011). State of the Art Inflation Targeting. Center for Central Banking Studies Handbook, No. 29, Bank of England, London.

• Horrigan M. W. (2013). Big Data: A Perspective from the BLS. Amstat News, January 1. Available at http://magazine.amstat.org/ blog/ 2013/ 01/ 01/ sci-policy-jan2013/

• Jahan S. (2017). Inflation Targeting: Holding the Line. Finance & Development (IMF). Available at http:// www.imf.org / external / pubs / ft / fandd / basics / target.htm

• Krsinich F. (2015). Price Indexes from Online Data Using the Fixed-Effects WindowSplice (FEWS) Index. Paper presented at the Ottawa Group, Tokyo, Japan, May 20-22, 2015.

• Nygaard R. (2015). The Use of Online Prices in the Norwegian Consumer Price Index. Paper prepared for the meeting of the Ottawa Group, Tokyo, Japan, May 20-22, 2015.

• Roger S. (2010). Inflation Targeting Turns 20. Finance & Development, Vol. 47, No. 1, IMF, pp. 46-49.

APPENDIX. TABLES AND FIGURES

Table 1. Descriptive statistics of official inflation

Index	% of CPI Basket	# of Sub-Indexes	S. D.	Min	Mean	Max
CPI	100	328	1.00	-0.36	1.04	3.52
Food	39.6	113	1.26	-1.06	0.81	3.42
- Bread	7.3	21	0.69	-1.15	0.66	1.96
- Meat	10.1	23	1.61	-1.32	1.33	5.15
- Fish	2.2	9	0.59	-0.90	0.15	1.26
- Milk	6.2	14	3.83	-5.15	1.29	11.43
- Fats	4.6	6	1.13	-0.36	1.16	3.60
- Fruits	2.3	10	5.46	-5.70	1.05	13.45
- Vegetables	2.4	16	10.04	-21.50	-0.31	16.69
- Sugar	3.4	7	0.92	-1.90	0.36	2.21
Beverages	1.4	7	0.28	-0.04	0.42	1.03
Alcohol	9.2	12	1.28	-1.91	1.69	3.33

Table 2. Descriptive statistics of online dataset

Index	% of CPI Basket	Rel. Share	# of Sub-Indexes	# of Goods, 1k	# of obs., 1m	Mean S.D.
CPI	45.7	45.7	134	75.1	2.48	4.96
Food	34.1	86.2	93	34.3	1.11	5.57
- Bread	6.8	93.0	19	8.19	0.29	3.39
- Meat	7.1	70.7	16	3.51	0.11	3.42
- Fish	2.2	100	9	2.57	0.09	4.01
- Milk	5.4	86.0	11	4.96	0.15	3.80
- Fats	4.4	96.8	4	0.72	0.03	3.99
- Fruits	1.4	60.2	5	0.74	0.02	9.66
- Vegetables	2.3	99.0	15	1.90	0.06	13.18
- Sugar	3.4	100	7	6.93	0.20	3.41
Beverages	1.4	98.9	6	10.5	0.40	4.71
Alcohol	6.2	67.6	7	9.92	0.37	3.59

**Figure 1. Selected online sub-indexes together
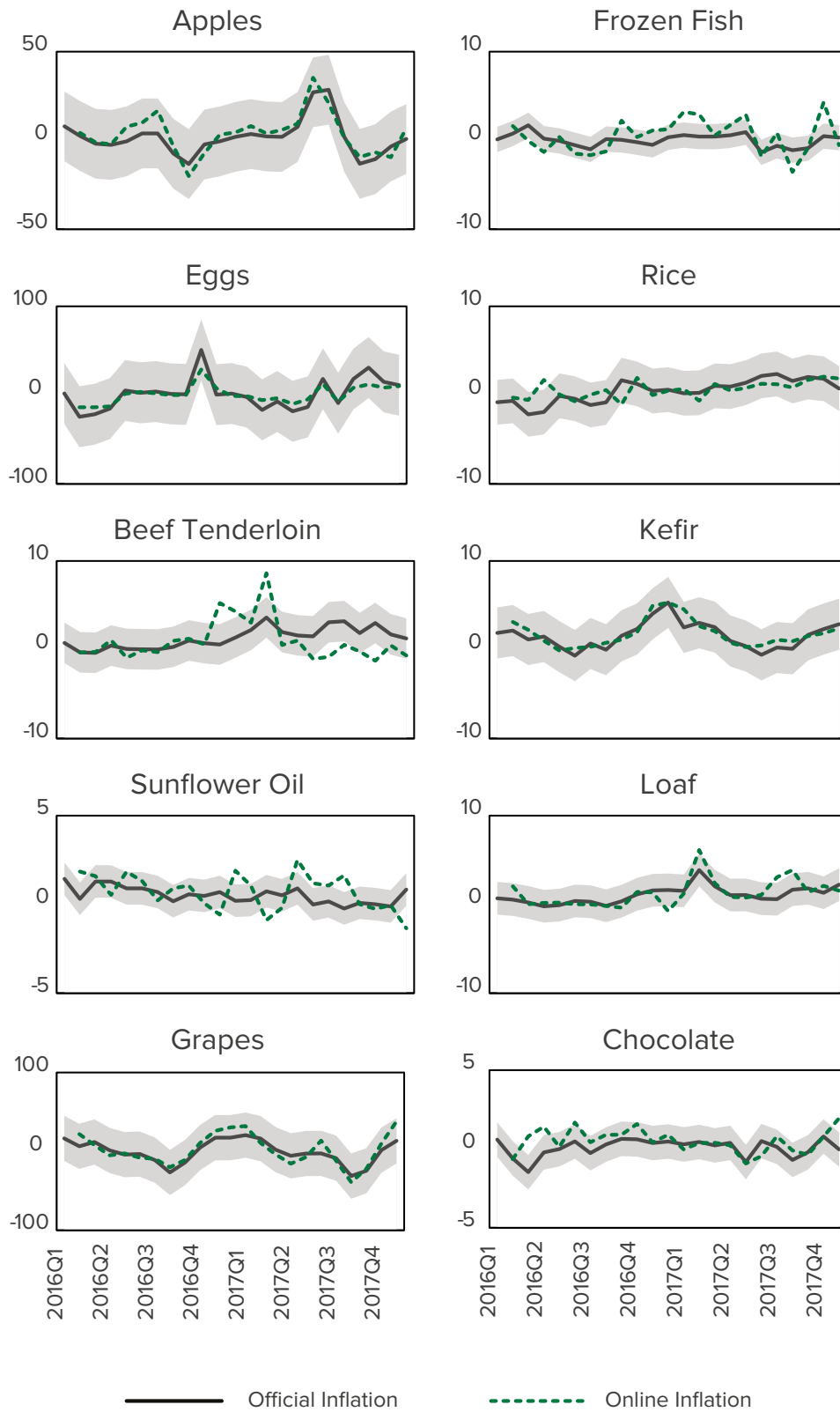with their officially provided counterparts, m-o-m %**

Official Inflation   Online Inflation

Table 3. Matching performance of online sub-indexes

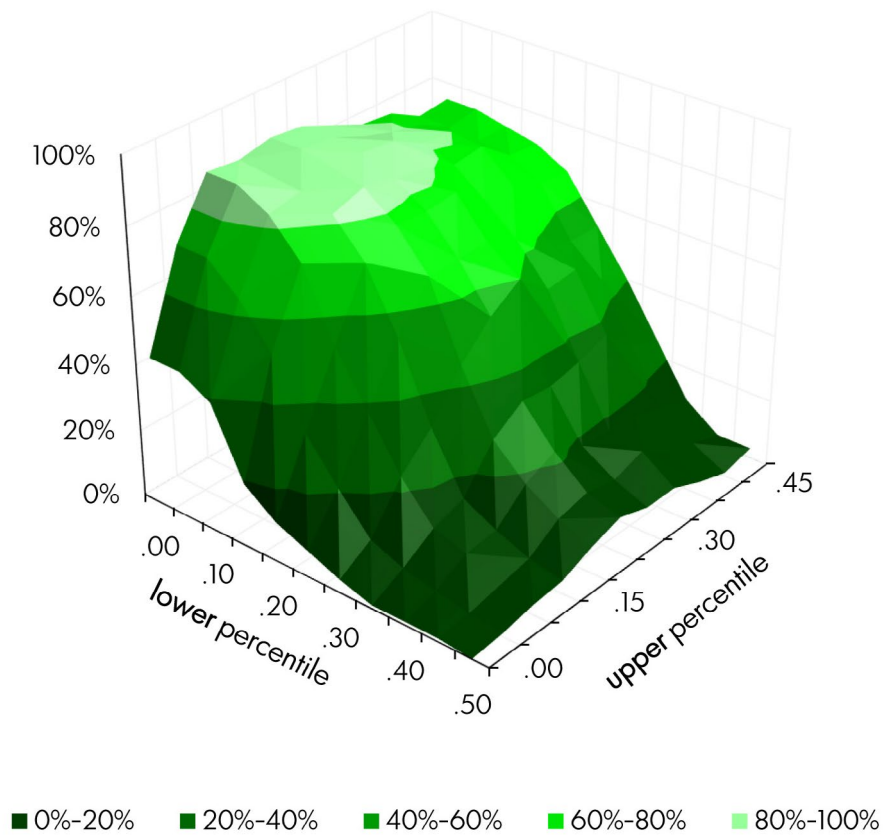|		Week I	Week II	Week III	Week IV
RMSE	*mean*	4.97	4.97	5.09	5.30
	min	0.73	0.78	0.65	0.78
	max	84.75	110.30	75.82	86.17
RMSE adjusted*	*mean*	2.68	2.54	2.61	2.73
	min	0.49	0.48	0.39	0.50
	max	11.13	11.70	11.76	28.08
Mean Overshoot Error		3.99	3.81	3.95	4.11
Mean Undershoot Error		2.89	2.67	2.83	3.14
Share of predominant overshoot errors		70%	72%	72%	72%

* RMSEs here are adjusted to the specific category's standard deviation of official inflation for comparison purposes.

Table 4. Determinants of online inflation performance (panel regression)
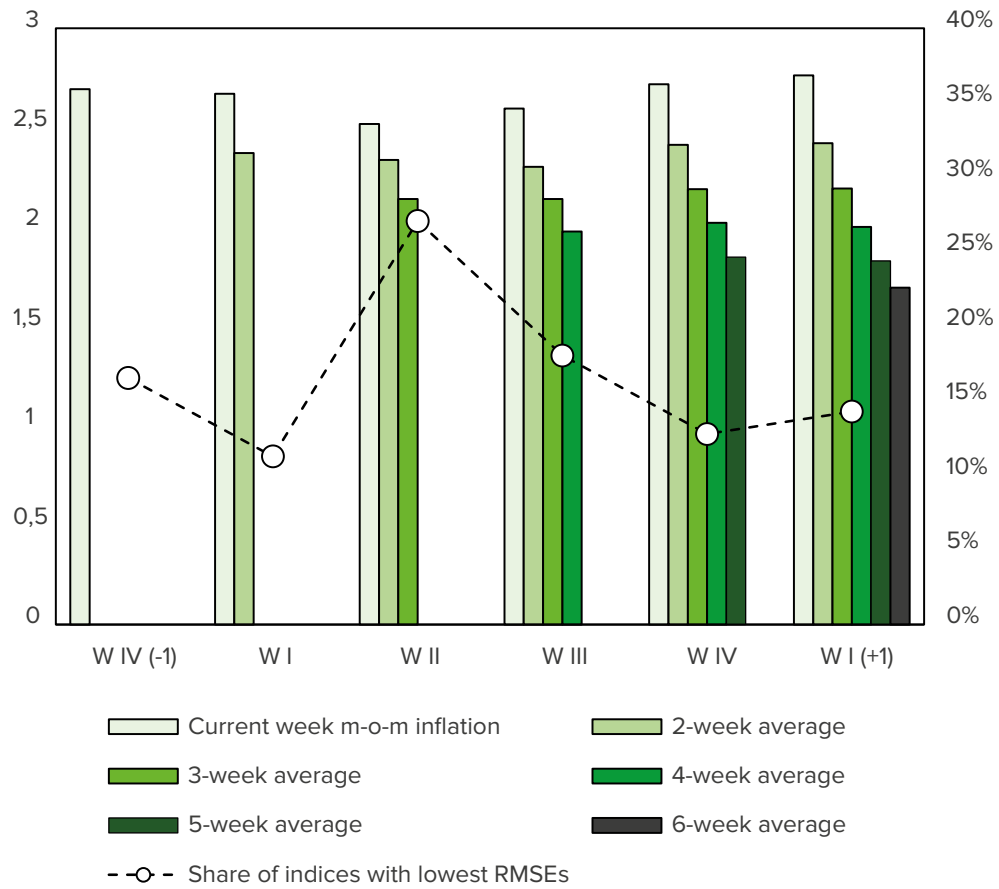
RMSE	1)	2)	3)	4)	5)	6)	7)
Average number of observations	8.661* (0.137)			8.470* (0.137)	3.957* (0.073)		3.784* (0.073)
Number of goods		-0.001* (0.000)		-0.001* (0.000)		-0.001* (0.000)	-0.001* (0.000)
Group's mean S.D.			1.870* (0.001)		1.868* (0.001)	1.870* (0.001)	1.867* (0.001)
Fixed effect	V	V	V	V	V	V	V
R²	0.844	0.838	0.955	0.844	0.955	0.955	0.956

Note: « * » indicates 1% significance level.

**Figure 2. Filtering-out lower and upper percentiles of inflation SDs:
Share of CPI sub-indexes with RMSE below average**

**Figure 3. Weekly performance of online indexes: Mean RMSE (ls),
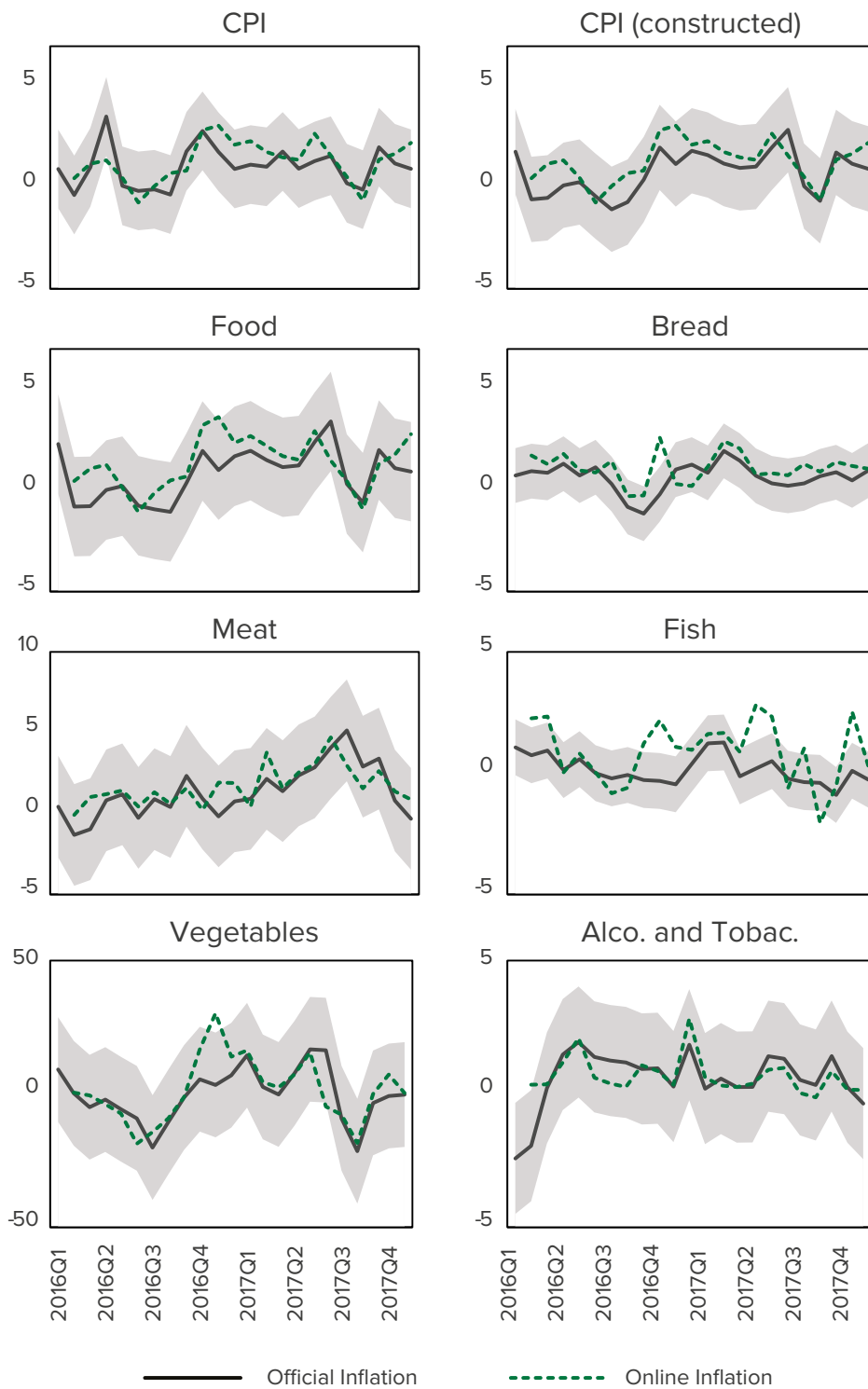Share of sub-indexes with lowest RMSEs (rs)**

Note: RMSEs here are adjusted to the specific category's standard deviation of official inflation for comparison purposes

Table 5. RMSE of aggregated month-on-month online indexes

Index	Simple average of all goods	Simple average of online indexes	Weighted average of online indexes
CPI	**0.81 (0.82)**	1.89 (1.90)	0.90 (0.90)
CPI (constructed)	1.06 (0.99)	1.92 (1.78)	**0.93 (0.87)**
Food	1.21 (0.97)	1.98 (1.58)	**1.14 (0.91)**
- Bread	**0.79 (1.14)**	2.31 (3.33)	0.85 (1.23)
- Meat	1.37 (0.85)	2.96 (1.84)	**1.07 (0.67)**
- Fish	**1.16 (1.97)**	2.60 (4.43)	1.33 (2.26)
- Milk	2.95 (0.77)	3.37 (0.88)	**1.61 (0.42)**
- Fats	4.65 (4.11)	**1.68 (1.48)**	**2.79 (2.47)**
- Fruits	**4.06 (0.74)**	5.85 (1.07)	5.43 (1.00)
- Vegetables	**8.67 (0.86)**	9.04 (0.90)	9.45 (0.94)
- Sugar	1.11 (1.21)	2.56 (2.79)	**0.78 (0.85)**
Beverages	0.81 (2.93)	2.47 (8.90)	**0.80 (2.90)**
Alcohol	1.20 (0.94)	2.51 (1.97)	**0.82 (0.64)**

Note: numbers in brackets represent RMSEs adjusted to the specific category's standard deviation of official inflation for comparison purposes.

Figure 4. Aggregated online and official infalation, m-o-m, %.

CPI

CPI (constructed)

Food

Bread

Meat

Fish

Vegetables

Alco. and Tobac.

——— Official Inflation   - - - - - - Online Inflation

Note: Aggregated web indexes here are constructed as a weighted average of sub-indexes which, in turn, are averages of four month-on-month web indexes