12th biennial IFC Conference: "Statistics and beyond: new data for decision making in central banks"

22-23 August 2024

Inflation and price dynamics analysis using high-frequency data from supermarkets: evidence from Peru [1]

Gonzalo Bueno,
Central Reserve Bank of Peru

[1] This contribution was prepared for the conference. The views expressed in this publication are those of the authors and do not necessarily represent the official views of the Committee, its members, or the BIS.

Inflation and Price Dynamics Analysis using High-Frequency Data from Supermarkets

Evidence from Peru

Gonzalo Bueno

I analyzed daily supermarket price data from Peru spanning the period of inflation increase from 2020 to 2022, followed by its subsequent reduction from 2023 and early 2024. The Central Reserve Bank of Peru lacked high-frequency price indices to monitor inflation and understand price formation. To address this gap, daily data from web pages of supermarkets was extracted and classified according to the COICOP classification. The findings reveal that price changes peaked later than official data but declined more rapidly. Price changes increased as inflation grew. This resulted in a negative correlation between inflation and duration of prices, which could be regarded as a shift in the inflation regime. Changes seem to be balanced between positive and negative values or biased towards the positive ones, so that variations related to trend prices are larger for increases than decreases.

Keywords: web scraping, inflation, retail prices, price duration

JEL classification: L11, L81, E31, C55

Contents

1. Introduction

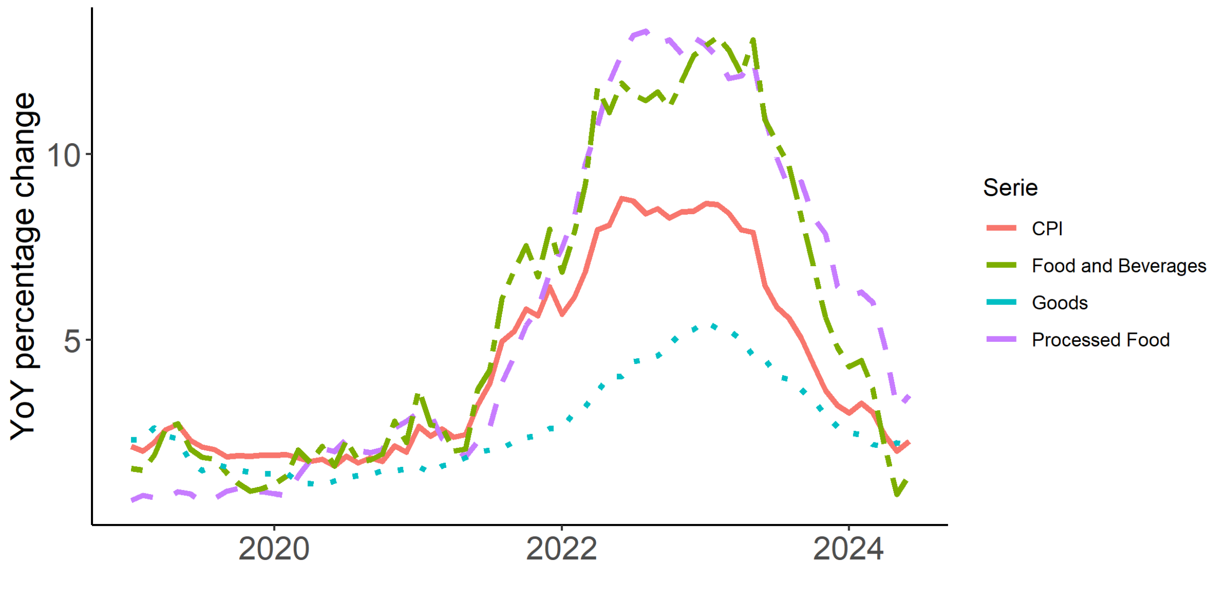
The inflationary process in Peru since the onset of the millennium has been characterized by a stable and low rate, akin to that of developed countries[1]. However, numerous supply shocks have occasionally caused the inflation rate to temporarily rise above its usual low level.

Sector-specific or idiosyncratic supply shocks were observed. For example, fuel prices affected transportation costs, indirectly raising the price of other goods and food items. On the other hand, a surprise shortage of fertilizers led to a rise in the price of some food items. Additionally, the depreciation of the domestic currency, having a dual role of increasing the cost of final and intermediate imported goods, added further pressure on prices. Finally, a steady increase in prices directly caused by these shocks led to higher inflation rates as they spread among intermediate inputs and final goods.

Such inflationary shocks were a major factor in the inflation surge in Peru observed since mid-2021 (See Graph 1), peaking in January 2023 (8.66 percent year-on-year). Food and beverages items had a persistent price increase from mid-2022. For goods items inflation (part of the trend component), a similar behavior was observed. Such episodes require a thorough monitoring of different CPI basket components.

YoY inflation rose since the second half of 2021 — Graph 1

A. Year-on-year percentage change



Sources: INEI

The analysis of headline inflation can be enriched by studying the process of individual price formation as detailed in Cavallo & Rigobon (2016). This is enhanced when information availability is at high frequency and not just monthly. Currently, the Central Reserve Bank of Peru (BCRP) lacks complete high-frequency data on prices of processed food and goods. In periods of low inflation, these prices remain relatively

[1] Around 2.7% from January 2000 to December 2019.

stable, making their weekly monitoring less critical for the monetary policy process. Nevertheless, during a period of high inflation, it is important to exploit all available information about the sources of price increases.

Therefore, to have a better understanding of the effect of inflationary shocks, deeper knowledge about the price formation process is required. A high-frequency price database allows us to observe shock propagation and changes in duration of prices. Having a database at the specific variety level (for instance, one-liter vegetable oil of a specific brand) can further enrich the analysis.

Hence, this study fills this statistical information gap by presenting a methodology for extraction, classification, and processing of high-frequency supermarket price data. Furthermore, it presents the analysis of collected prices, their year-on-year trend, the frequency of change and duration and how it has evolved in the recent years during the increase in inflation.

Related Literature

Research on the use and construction of high-frequency inflation indicators is relatively recent and scarce. Typically, studies have focused on building high-frequency inflation indicators, making monthly inflation predictions from daily or weekly information, and analyzing aspects such as the duration or frequency of price changes to measure the degree of price rigidities.

An initial study on the use of internet data is Lunnemann & Wintr (2006) which compares price rigidities in-store versus online prices. This referred study was even published before the creation of the Billion Prices Project (BPP) in 2008 (see Cavallo & Rigobon (2016)), which involved extracting online prices from stores in up to 50 countries. Although the BPP has been discontinued, it gave rise to PriceStats, a private entity that collects and sells price information and high-frequency inflation indicators for several countries. It is worth noting that there is no price information for Peru in either BPP or PriceStats.

With information from BPP and PriceStats, various analytical studies have been developed. For instance, Cavallo (2013) found that the measure of inflation based on online prices is similar to the official general inflation in Latin America, except in Argentina. Meanwhile, Aparicio & Bertolotto (2020) analyze the utility of data extracted from the web from July 2008 to September 2016 in 10 advanced countries to forecast inflation. Recently, Macias, Stelmasiak, & Szafranek (2022) use data obtained through web scraping at the National Bank of Poland to conduct inflation nowcasting experiments to show that the use of these data improves the accuracy of inflation projections.

Furthermore, the duration or rigidities of online prices are studied, for example, in Gorodnichenko & Talavera (2017) or Cavallo (2018). Specifically, Cavallo (2018) finds that online prices in the U.S. for non-food products became more flexible from 2014, reaching levels similar to those of food prices. In the case of Peru, (Coronado, Lahura, & Vega (2020) conducted a preliminary study on online price rigidities.

A major challenge in using web scraped data in the analysis of high-frequency prices is the classification problem. Given the large volume of unstructured data, the use of specialized machine learning algorithms is inevitable. Consequently, a substantial body of literature has emerged, primarily from the data science community.

The paper proceeds as follows, section 2 describes the data extraction method. Section 3 describes the product classification methodology according to the

Classification of Individual Consumption According to Purpose (COICOP). Section 4 deals with price filters applied for data analysis. Section 5 presents the results and the analysis of price behavior. Finally, section 6 concludes and sketches the pending research agenda.

2. Data Extraction Methodology

Using web scraping techniques, data on prices was collected for all product varieties listed on the websites of two supermarkets since April 2020. The algorithm retrieves basic information from each product listing on the web such as: (i) name, (ii) price, (iii) brand, and (iv) category.

Approximately, 134,000 unique product variety records that have been sold at any time were obtained, regardless of whether they are currently available. The main reasons for a record ceasing to update include the discontinuation of its sale or a change in its presentation. For instance, companies recover costs without altering prices by reducing the weight of their product presentations, which results in a new variety of the product.

On supermarket websites, products can have up to three prices: (i) the regular price, (ii) the "online" price or discounted price, and (iii) the price with a special discount using a specific card. The online price is always available if the product listing is present on the web, whereas the regular price only appears when there is a discount for purchasing the product online. When there is no discount, both prices match. The price used is the discounted one (or the regular one if it is not available), which is the price a consumer would pay.

Each product is differentiated by its name. A change in it generates a new price series. For example, a wine initially registered as "Vino rosé" ("Rosé wine"), with a tilde in its name, and later as "Vino rose" ("Rose wine"), would enter the database as two distinct products. Similarly, if one has "Pack galleta de chocolate" ("Pack chocolate cookie") and "Galleta de chocolate pack" ("Chocolate cookie pack"). To avoid this, each product was assigned an identifier name consisting of its original name with a cleanup of unusual characters (such as tildes or hyphens) and without capital letters. Nonetheless, the latter case cannot be avoided with this cleanup and requires an algorithm that can find similar names in the database. This has not yet been implemented.

3. Product Classification Methodology

The downloaded series are associated with a classification according to the supermarkets' criteria. As the goal is to have new sources of information for monitoring inflation, it is necessary to have a grouping according to the CPI categories. These correspond to the COICOP. Thus, a classification is constructed over a basket of food, beverages, and goods grouped into twelve divisions and disaggregated to the level of specific variety.

Divisions of the CPI basket and their levels

Graph 2

Level 1	**Divisions**	Food and non-alcoholic beverages
Level 2	**Groups**	Food
Level 3	**Classes**	Bread and cereals
Level 4	**Subclasses**	Rice in all forms
Level 5	**Category**	Rice
Level 6	**Product**	Packaged Rice
Level 7	**Item**	"Extra" Packaged Rice
Level 8	**Specific variety**	Paisana rice orange bag 1kg

Sources: INEI

The available data from INEI (National Institute of Statistics and Informatics) shows the categories of food and beverages at the item level, and for goods, at the subclass level. To gain a better understanding of the products and varieties that make up the CPI basket, the classifications were reconstructed to the variety level with the information available in the National Household Budget Survey (ENAPREF). This survey records household expenditures across the divisions. With the COICOP identification code at hand, the different categories by varieties can be recognized.

Supermarket products were classified using machine learning techniques, with a supervised learning model for text classification (Joulin, et al., 2017). In this case, each specific variety name was associated with the corresponding product type. For example, “ABC brand corn oil 1L” is classified as “packaged vegetable oil”. This procedure required two stages: (i) cleaning the text identifier name to make it suitable for classification and (ii) creating a training sample.

The first stage involved eliminating units of measure as they do not provide relevant information about the product type. Thus, a product like “ABC brand corn oil 1L” would be changed to “corn oil ABC brand”. An outstanding issue is the identification of brands to remove them to improve inputs to the algorithm.

For the second stage, out of approximately 22,000 records of food and beverages from one supermarket, 50 percent were manually classified to build a training sample, which was then used to classify the remaining 50 percent, as well as data from other supermarkets. A similar procedure was done with a sample of 16 thousand goods.

4. Price Series Filters

Data Imputation

Initially, the extraction code was manually executed at three days a week frequency. However, data was unavailable for some weekends. Since 2022, an automated routine was adopted to address this issue. Given this situation, an algorithm was employed to fill in missing data for some days. For this purpose, the duration of the gaps without data is examined, and if they are less or equal than thirty days, it is filled with the last price (e.g. a product with a price of 2 soles [Peruvian currency] on Monday, which has no data for one week and then reappears on the following Monday).

30-day moving average

To smooth the price series, a thirty-day moving average was calculated with at least 7 days of data. That is, if out of the last 30 days at least 7 days have data, the mean is calculated.

Price series exhibit volatile behavior due to the presence of short-term price discounts and increases. Graphically, one can plot a stable price around which the price fluctuates. The thirty-day moving average filter smooths out the series against these jumps, however, it still captures non-trend movements.

Regular Price Extraction

An alternative approach is to use a filter based on Nakamura & Steinsson (2008). This algorithm detects discounts or sales by identifying price patterns in symmetric and asymmetric V shapes. For example, in the first type, a product might have a value of 10 soles in January, then decrease to 8 soles (discount) in February and return to 10 soles in March. The pattern is symmetric because it returns to the value before the discount. In the second type, the rebound can be to higher values. For instance, in March, the price would be 12 instead of 10 (See Graph 3.A).

On the other hand, it may happen that discounts persist for a temporal window of such length that it could be assumed that the price level has been reduced. Therefore, it is necessary to establish a limit to that window corresponding to the maximum assumed duration of a discount.

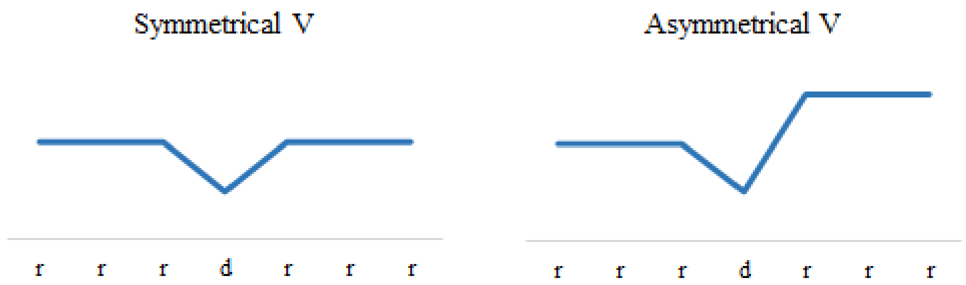
Furthermore, during processing, an additional problem was found: the simultaneous presence of a price discount and a drop in the regular price. For example, a price that drops from 10 soles to 5 soles for a few days, and then remains constant at 6 soles. The middle price should be considered as a discount if it does not exceed a time limit (See Graph 3.B).

Due to the exploration of differentiated behavior within food items, the need to establish different time window limits in the described algorithm is evident. Otherwise, long discounts could be confused as price changes. Since regular and discounted prices are extracted, the duration of discounts can be measured. For food and beverages, the average time is around 30 days, while for goods, it ranges from 30 to 60 days approximately. Therefore, for this article, a maximum discount window of 30 days was established for all products (See Graph 3.B).

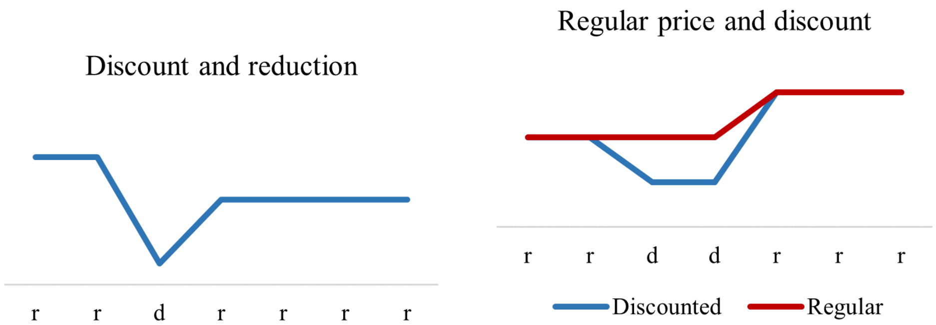
After filtering discounts, price series with short-lived upward jumps were noticed. The filter was adapted to detect these phenomena, and if they last less than a previously established window, they are not considered as an increase in the regular price. For this article, this value was set to 7 days (See Graph 3.C).

The filter smooths prices to capture trend movements Graph 3

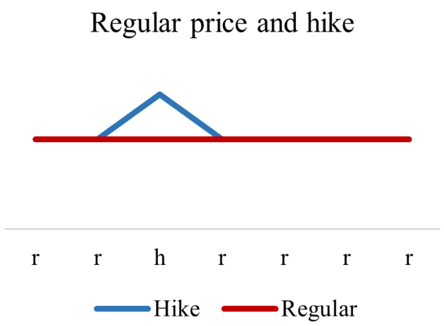
A. Difference between symmetrical and asymmetrical V shape



B. The filter also recognises reductions after discounts and fills those periods



C. Price hikes are also smoothed



Date axis	
r	Day with regular price
d	Day with discounted price
h	Day with hiked price

Sources: Nakamura & Steinsson (2008) and BCRP.

As a result, the proposed algorithm requires of a vector of three parameters: the first element of these vectors control how staggering filtered prices will be. The higher this parameter, the filtered price series becomes staggered and monotonically increasing, as more price reductions are filtered out because they are considered as discounts. On the other hand, if this parameter approaches zero, the filtered series becomes similar or equal to the original series. The second element controls the maximum duration allowed for discounts before price reductions. Finally, the third element captures the minimum duration of a price increase to consider it a trend movement rather than a data mistake or outlier.

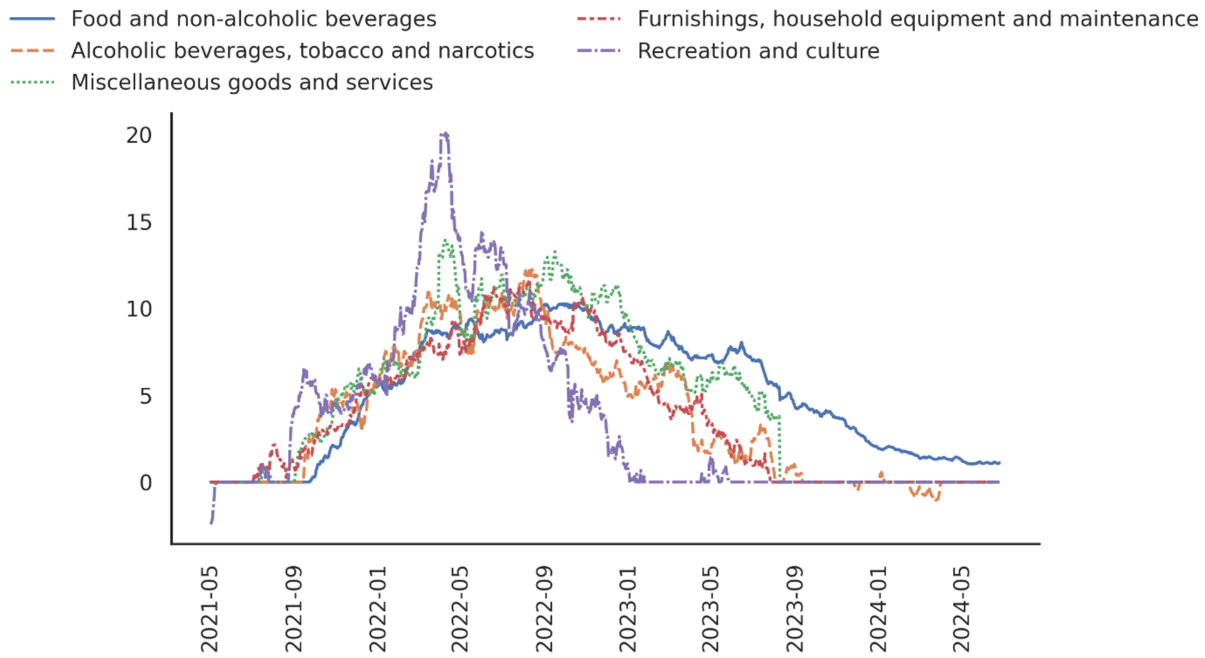
5. Results

Moving average series

I estimated median values of year-on-year percentage variations using price series produced with the moving average filter (See Graph 4). The increase in inflation intensified during 2022. However, prices of food and beverage items differ from those of other type of goods due to their persistence. Other goods prices experienced higher increases, but since early 2023, they moderate. For the three types of good items in Graph 4, median price changes became anchored at zero at the end of the sample. This evidence points towards a differentiation of low and high inflation regimes. In the first one, the median of price changes is equal to zero, and it becomes positive as the regime change and shocks trigger an increase in the proportion of products that update their prices.

Price increase indicators peaked in the second half of 2022 — Graph 4

A. Median of yoy percentage change



Sources: Supermarket data, BCRP.

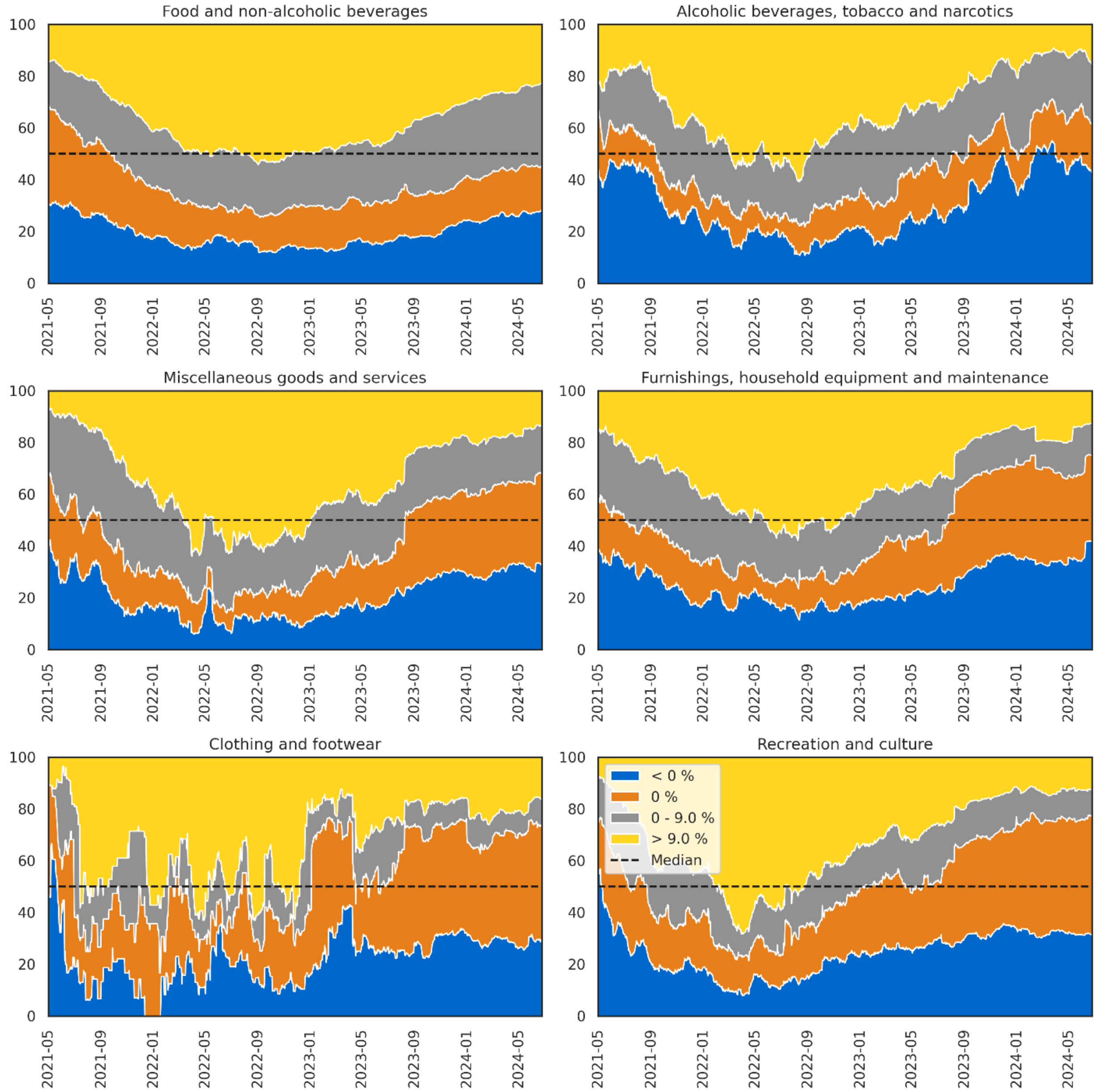
Graph 5 provides a distribution of price changes according to ranges in each of the divisions analyzed. The proportion of food items with price increases above 9 percent year-on-year grows and diminishes slowly barely reaching 50 percent by October 2022. Two good items divisions show a sharper change in the distribution.

Also, the proportion of prices with zero growth increases faster by the end of the sample in the goods category. Since the filter used is a moving average, discounts in prices can generate variations in the series such that the percentage variation is different from zero exactly. Thus, values that are equal to zero coincide with prices that have not varied over a year considering discounts.

In the second half of 2022, annual increases in prices peaked and by the end of 2023 they slowed down. Food and beverages prices moderate smoothly

Graph 5

A. Year-on-year percentage change distribution areas



Sources: Supermarket data, BCRP.

By choosing subsamples within the food category (see Graph 6), differentiated behaviors and shocks are highlighted. The four groups are Non-Durable Agricultural Foods, Foods Dependent on International Prices, Processed Foods, and Other Non-Durable Foods.

Prices of Perishable Agricultural Foods react to changes in the market supplies associated with climate conditions and social turmoil related to blockage of roads that prevent transportation of products. Also, the increase in fertilizer prices affected the productivity of crops.

The sample starts with a decrease in the median in May 2021, that coincides with the reduction in official data due to an increase in market supply. After the normalization of the year-on-year variation, the median is equal to zero. After March 2022 the official data shows the beginning of an upward trend, caused by an increase in rain and obstruction of roads, that coincides with a positive median for the online data. By the end of 2022, the median becomes stable, as the official year-on-year variation peaks due to the fertilizer shortage[2]. The daily data seem to capture how shocks affect individual prices, but the rate of adjustment differs from the monthly index.

Foods Dependent on International Prices faced supply shocks associated with fright costs, commodity prices and the increase in the nominal exchange rate. Official data showed a peak in April 2022 that coincided with the first peak in supermarket data. It seemed that both series captured the increase at the same rate. Nonetheless, the official index change trend decreased by June 2023, while the supermarket prices showed a decline after August 2023. This may be due to a difference in statistical methods but also could be associated with asymmetric price adjustments at the micro level.

Processed foods are characterized by less volatile prices. From September 2021, the median of the year-on-year percentage changes rose above the zero threshold. It peaked on the first half of 2022 and follows a steady trend with shows less abrupt changes compared to the other samples. It resembles how core inflation behave in official data.

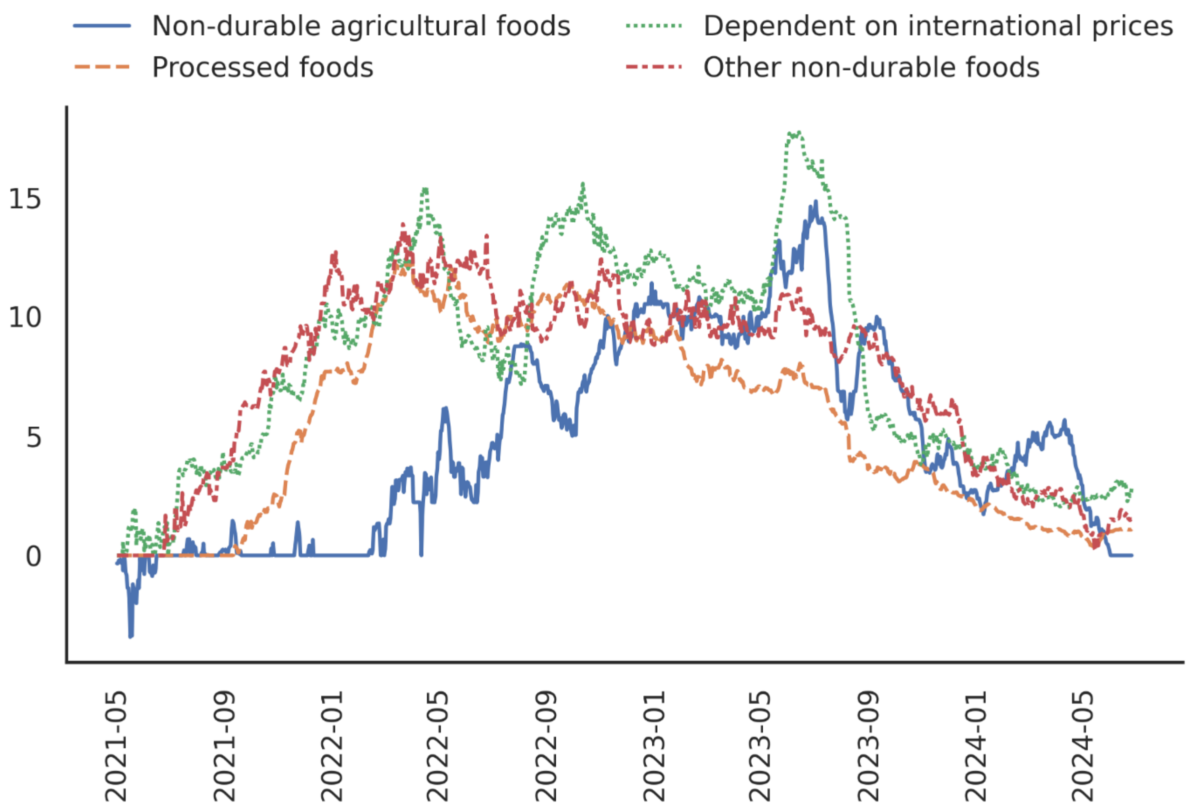
Finally, Other Non-Durable Foods price variations seem to follow a common trend with two samples, but in official data, it is characterized by a volatile and particular path during some periods.

[2] Because the analysis is on the variation of raw prices and no index is being calculated, year-on-year percentage changes behave differently. Official data smoothly reach peaks and then slowly reverts to common values. Supermarket products will have abrupt percentage changes when a price goes up, and after 365 days, it will go back to zero (conditional to showing no further changes).

Idiosyncratic shocks made agricultural foods take a different path of year-on-year variation while processed foods seem to follow a stable trend

Graph 6

A. Median of yoy percentage change



Sources: Supermarket data, BCRP.

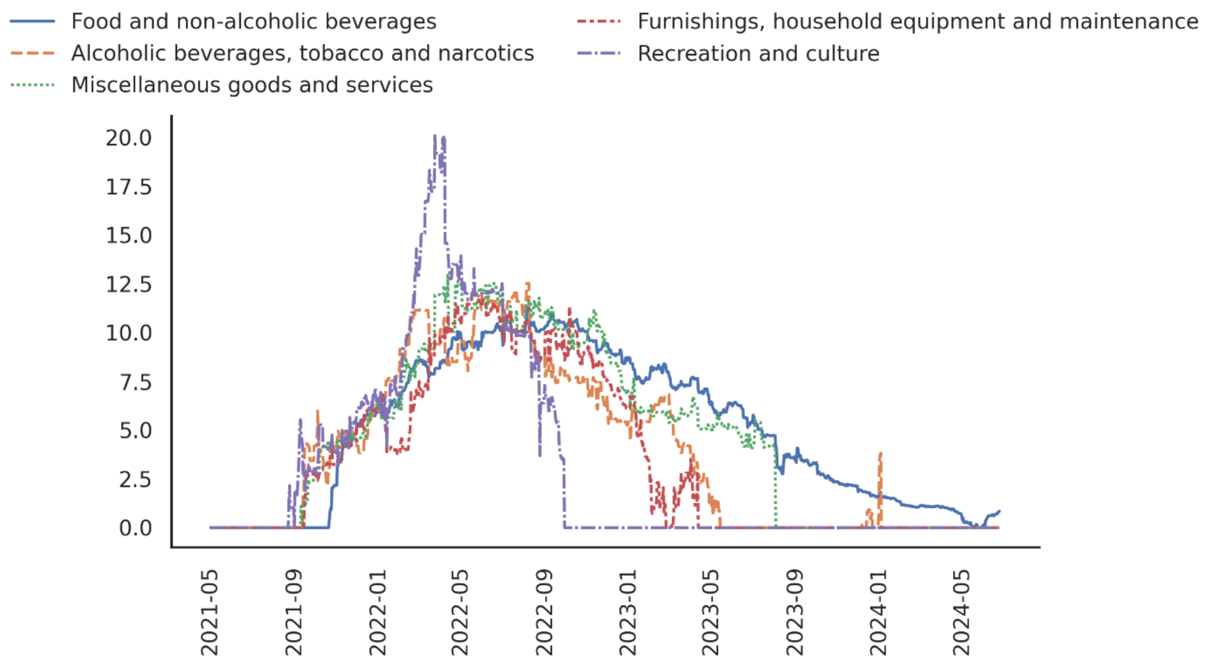
Series with regular price filter

I replicate the figures of median values and distributions estimated with the regular price series. The results about trends, the price surge of 2022 and subsequent moderation, remain (See Graphs 7, 8 and 9). This implies that the trend of price changes can be detected even with the presence of temporal price reductions.

The inflation trend with regular filtered prices coincides with the smoothed moving average filter

Graph 7

A. Median of yoy percentage change



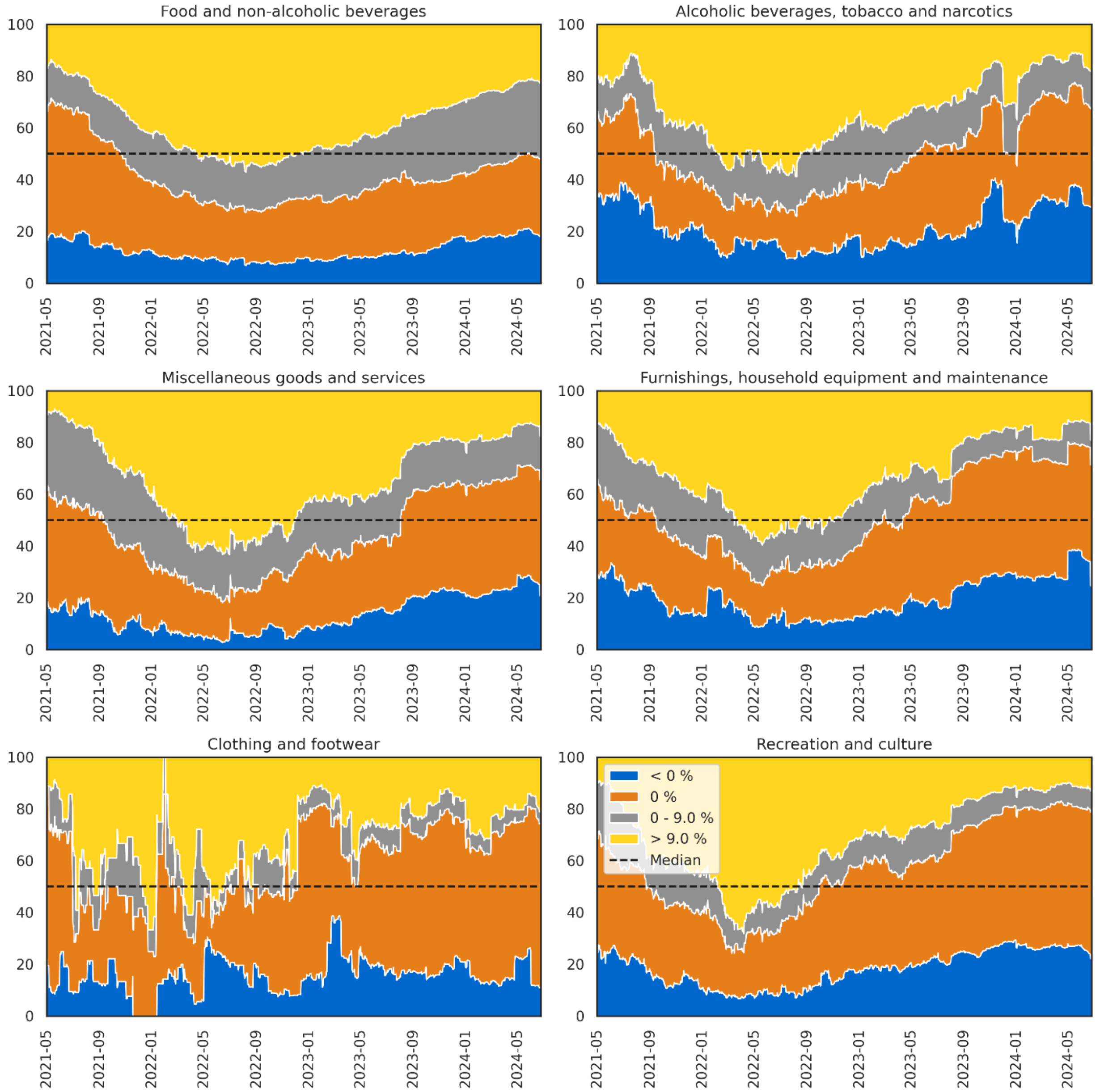
Sources: Supermarket data, BCRP.

The regular price filter captures sharper changes unlike the smoothing of the moving average. As a result, the proportion of products with zero variation is higher. Furthermore, the removal of discounts reduces the proportion of price decreases. Nonetheless, regular prices seem to follow the same evolution in distributions as captured with the moving average, and the swap between the yellow and orange areas points towards two different year-on-year behavior, as more products update and change prices. This result adds to the differentiation between a high and low inflation regime.

Although the proportion of products with without year-on-year price changes increases, the trend of inflation and moderation holds

Graph 8

A. Year-on-year percentage change distribution areas

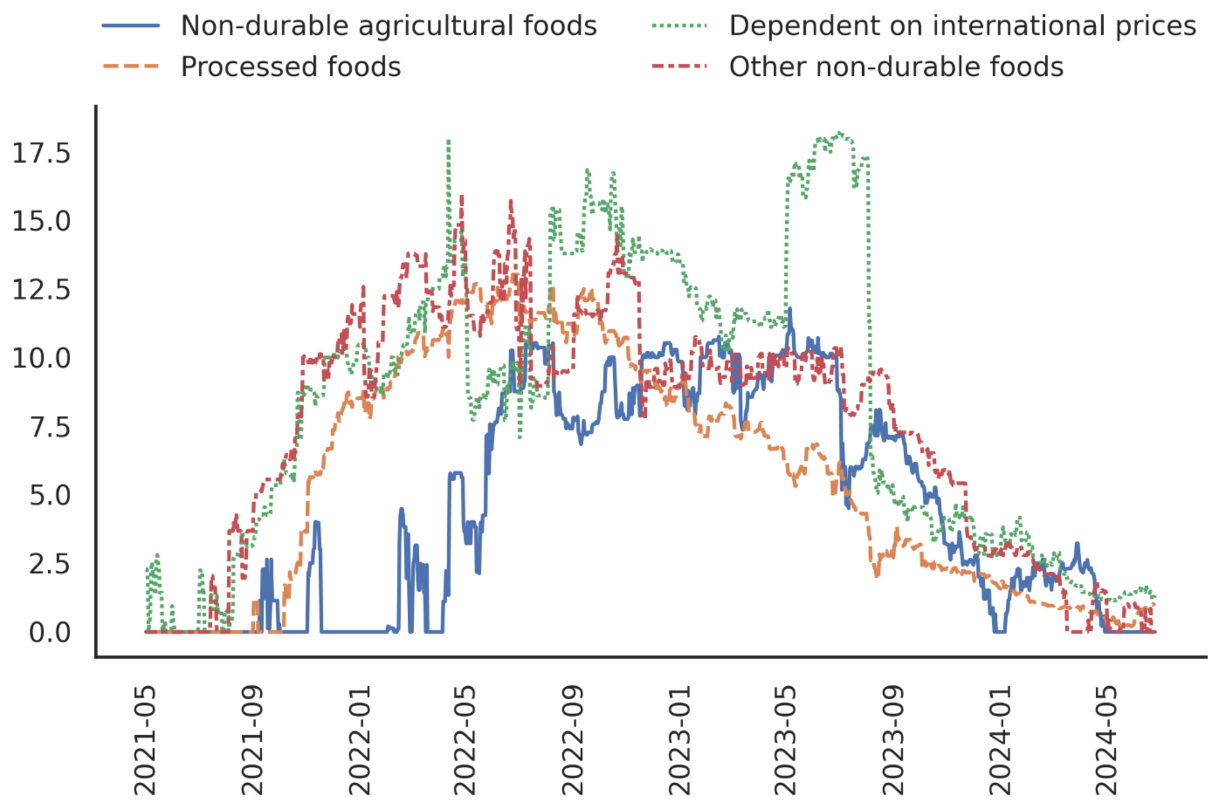


Sources: Supermarket data, BCRP.

Elimination of temporal price reductions does not make a big difference in the trends of foods prices

Graph 9

A. Median of yoy percentage change



Sources: Supermarket data, BCRP.

Comparison of Filters and Robustness Testing

To gauge the robustness of the filters, I assume two additional parameterizations of the filters given by the parameter vectors (100, 30, 30) and (7, 7, 1). They are mainly differentiated by the first parameter, that regulates how long does it allow discount to last.

The median series on year-on-year variation do not differ in trend and direction when comparing the moving average to regular price filtered series (see Graph 10). Three results stand out.

The series with the 7-day filter show abrupt increases and reductions. This is because it is the parameterization with less smoothing, and price changes can be noisier. Nonetheless, there are products, such as some foods, whose variation is so frequent that this parameterization could be more convenient for capturing the price direction without losing information on short-term volatility.

For foods dependent on international prices, the 100-day filter stands out for deviating from the common trend in mid-2022 and 2023. For this case, a reduction in prices in 2022 was filtered by the 100-day parameterization. As a result, in that year, prices did not decrease, moderating the median for the same months in 2023.

Finally, in the face of a permanent increase in prices, the moving average takes 30 days to update completely. That is part of its smoothing component. However, the stable price filter does not necessarily have this problem. This can be noted in some sections of the graphs where the moving average curve is slightly shifted to the right[3].

Detected inflation trends are robust to the specification of filters (including the Mov. Average). The 100 days filter returns the least volatile index

Graph 10

A. Median of yoy percentage change

Sources: Supermarket data, BCRP.

Price changes, frequency of change and duration

Available CPI data lacks information on price behavior and how they change and evolve. The constructed databased allows to study the size and number of changes, how frequent they are and how long do price levels last.

[3] Nevertheless, there is a limitation with this methodology. If a price undergoes a reduction that is not associated with a discount, the filter will take a number of days equal to the first parameter to differentiate the permanent change.

Table 1 shows that price changes are balanced between positive and negative and homogeneous across product divisions. Eliminating discounts does not affect this result. Food and Non-Alcoholic Beverages and Miscellaneous goods and services are skewed towards increases of prices. Prices could adjust upwards to generate inflation by having more positive than negative changes or by increasing more than the decreases.

Positive and negative price changes as proportion of total changes

Table 1

Division	Regular prices		30 days filter prices	
	Positive	Negative	Positive	Negative
Food and non-alcoholic beverages	65.7	34.3	65.6	34.4
Alcoholic beverages, tobacco, and narcotics	55.3	44.7	55.2	44.8
Clothing and footwear	49.5	50.5	49.6	50.4
Furnishings, household equipment and maintenance	51.9	48.1	51.7	48.3
Recreation, sport, and culture	49.1	50.9	48.8	51.2
Miscellaneous goods and services	61.9	38.1	61.9	38.1

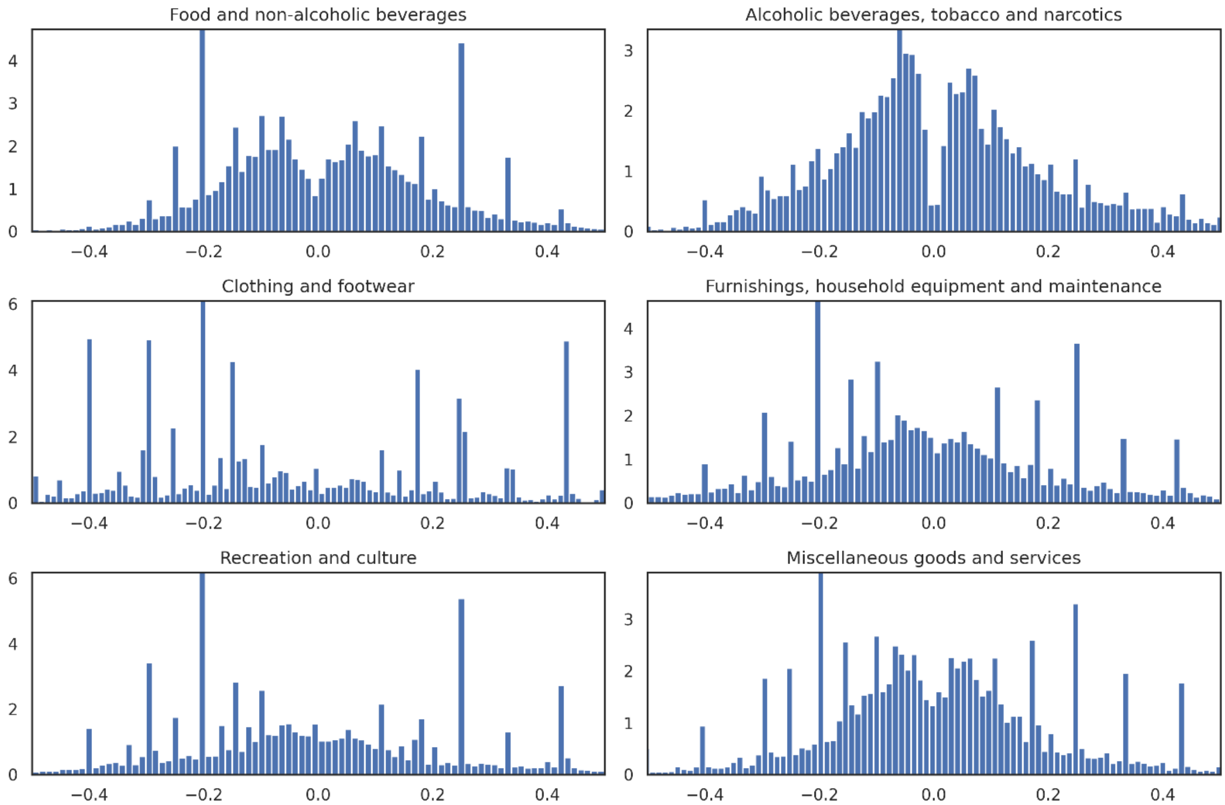
Sources: Supermarket data, BCRP.

Price changes follow a distribution with two peaks (see Graph 11), with a low proportion of variations close to zero. That is, it is more common to have price jumps than small reductions or increases. This finding is further supported by the high frequency of variations of +-20 percent, consistent with individual price literature.

However, this behavior is less pronounced for "furnishings, household equipment, and maintenance". In the case of "recreation and culture" and "clothing and footwear," the distribution is more uniform due to the higher presence of changes. Thus, it could be argued that adjustment behavior depends not only on the inflation regime but also on the type of product. Therefore, predictions from macroeconomic models should vary depending on the composition of the basket of goods.

Price percentual changes are bimodal, showing less density on small changes — Graph 11

A. Distribution of yoy percentage change



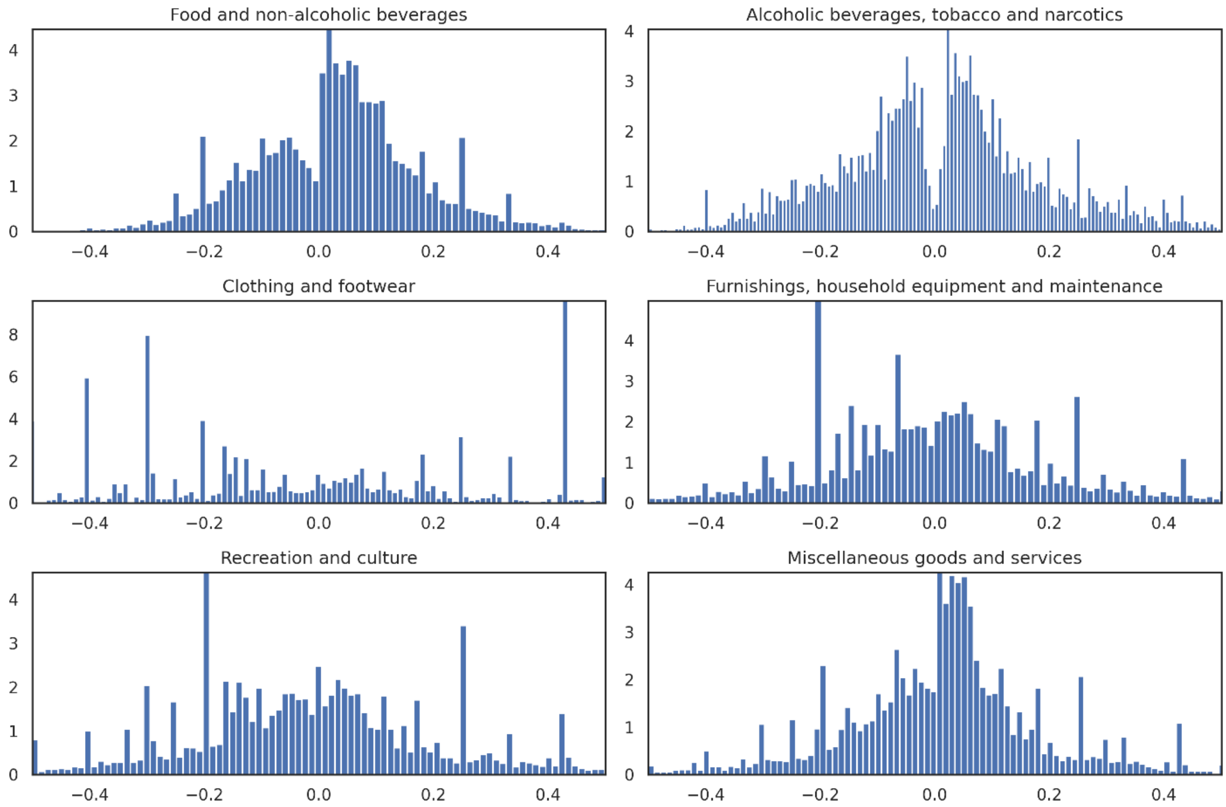
Sources: Supermarket data, BCRP.

Data with the 30-day filter show an asymmetric distribution (see Graph 15), with a greater presence of positive values. This may be due to the effect of discount elimination, providing a better perception of the frequency and magnitude of persistent price changes. Nonetheless, the sample covers the inflationary period after 2020, which may introduce bias.

Eliminating discounts show distributions of permanent changes that are biased towards more positive values

Graph 12

A. Distribution of yoy percentage change



Sources: Supermarket data, BCRP.

Frequency of price changes can be measured as the number of changes within a time period, and it can also be expressed as the fraction or percentage of time during which the product experiences changes. For instance, a product whose price changes 2 times in 90 days would have a frequency of 1/45, meaning it changes every 45 days, or the daily probability of change is 2.22 percent.

The duration of a price can be implicitly calculated from the frequency of price changes. The probability of a change occurring is subject to a distribution that may depend on time (see Nakamura & Steinsson (2008) and Kiefer (1988)). It is assumed that price changes follow an exponential distribution fit $f_{it} = 1 - \exp(-\lambda_i t)$where λ_i is the hazard function or conditional probability of change at $t + 1$ given that the price has survived until t. This distribution has the property of having a constant hazard function over time. Finally, if it is also assumed that the probability is uniform, then $\lambda_i = -ln(1 - f_i)$. The implicit duration is the inverse of this function: $d_i = \frac{1}{\lambda_i}$. In the previous example, if the monthly probability was 2.22 percent, the value of λ_i would be 44.5 days (a result similar to the duration calculated directly from the update frequency).

Prices without filtering exhibit a rapid update frequency, especially for food and non-alcoholic beverages (see Table 2). For this group, the implicit duration is less than a month, while for the rest of the samples, it ranges between 32 and 61 days. Nonetheless, increasing the time limit parameter for discounts reduces the frequency.

Assuming that discounts last no more than a week, the duration is increased accordingly. Conversely, with discounts of less than 30 days, the frequency is around a third. Generating time series with fewer price reductions, taking 100 days yields durations of approximately 5 months for food and beverages, while for goods, it ranges from 7 to 13 months. Another interpretation of this result is as a measure of the upward adjustment frequency, tolerating extended periods of price reduction.

Two-thirds of price changes are eliminated by removing discounts. Regular prices seem to last for five to thirteen months

Table 2

	Frequency of change (change probability)							
	Median				Mean			
Division	Raw	7	30	100	Raw	7	30	100
Food and non-alcoholic beverages	4,0%	2,3%	1,0%	0,6%	3,9%	2,0%	0,9%	0,6%
Alcoholic beverages, tobacco and narcotics	3,5%	1,9%	0,9%	0,4%	3,2%	1,9%	0,9%	0,4%
Clothing and footwear	2,4%	1,6%	0,6%	0,2%	2,5%	1,8%	0,6%	0,2%
Furnishings, household equipment and maintenance	2,0%	1,4%	0,7%	0,4%	2,4%	1,6%	0,7%	0,4%
Recreation, sport and culture	1,5%	1,2%	0,6%	0,4%	3,2%	2,1%	0,9%	0,5%
Miscellaneous goods and services	2,5%	1,7%	0,7%	0,4%	2,5%	1,6%	0,7%	0,4%

	Implicit duration (in days)							
	Median				Mean			
Division	Raw	7	30	100	Raw	7	30	100
Food and non-alcoholic beverages	24	44	104	154	25	49	109	166
Alcoholic beverages, tobacco and narcotics	28	53	115	224	31	52	116	228
Clothing and footwear	41	61	164	437	40	55	163	426
Furnishings, household equipment and maintenance	49	69	142	251	40	64	138	225
Recreation, sport and culture	66	86	167	277	31	48	109	196
Miscellaneous goods and services	39	59	144	237	40	60	146	243

Sources: Supermarket data, BCRP.

If we calculate the frequency for each quarter (see Table 3), we observe and increase during 2021 and 2022, followed by a reduction in 2023, for various filter specifications. Nevertheless, for raw prices, the frequency does not show that trend. This implies that after the increases in 2021 and 2022, there is a high number of discounts that are detected by filters that generate jumps in the series. The trend points towards a reduction in the frequency of changes that coincides with the shifts in the distribution of year-on-year variation and the moderation in inflation.

During the inflation increment, the duration of prices seems to have decreased, as more upward adjustments took place

Table 3

	Frequency of change (median)															
	Raw				7 days				30 days				100 days			
Division	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV
Food and non-alcoholic beverages	1.7%	4.0%	3.6%	4.1%	1.3%	2.9%	2.3%	2.4%	0.8%	1.3%	1.2%	1.1%	0.3%	0.9%	0.9%	0.6%
Alcoholic beverages, tobacco and narcotics	1.6%	2.1%	3.0%	2.7%	1.4%	1.6%	2.4%	1.4%	0.8%	0.6%	0.9%	0.5%	0.3%	0.3%	0.6%	0.2%
Clothing and footwear	0.0%	4.9%	1.6%	3.0%	0.0%	3.3%	1.1%	2.1%	0.0%	1.8%	0.6%	0.5%	0.0%	0.2%	0.2%	0.2%
Furnishings, household equipment and maintenance	1.7%	3.5%	3.2%	1.9%	1.2%	2.6%	2.2%	1.4%	0.6%	1.0%	1.0%	0.6%	0.3%	0.7%	0.7%	0.2%
Recreation, sport and culture	0.9%	1.3%	1.4%	1.5%	0.7%	1.1%	1.1%	1.1%	0.3%	0.2%	0.5%	0.6%	0.1%	0.1%	0.3%	0.2%
Miscellaneous goods and services	2.5%	2.4%	2.8%	2.9%	1.9%	1.9%	1.7%	1.9%	0.7%	0.9%	0.7%	0.7%	0.4%	0.7%	0.4%	0.4%

	Implicit duration in days (median)															
	Raw				7 days				30 days				100 days			
Division	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV
Food and non-alcoholic beverages	57	25	28	24	74	34	42	41	124	76	84	93	310	106	107	176
Alcoholic beverages, tobacco and narcotics	63	46	33	37	69	60	42	73	126	154	107	183	353	326	169	423
Clothing and footwear		20	60	33		29	89	46		54	159	196		636	460	665
Furnishings, household equipment and maintenance	58	28	31	52	84	39	45	69	155	104	97	171	298	136	145	578
Recreation, sport and culture	110	76	69	67	134	91	93	87	311	471	186	165	919	681	361	451
Miscellaneous goods and services	40	42	36	35	53	51	60	51	138	107	143	146	265	143	247	244

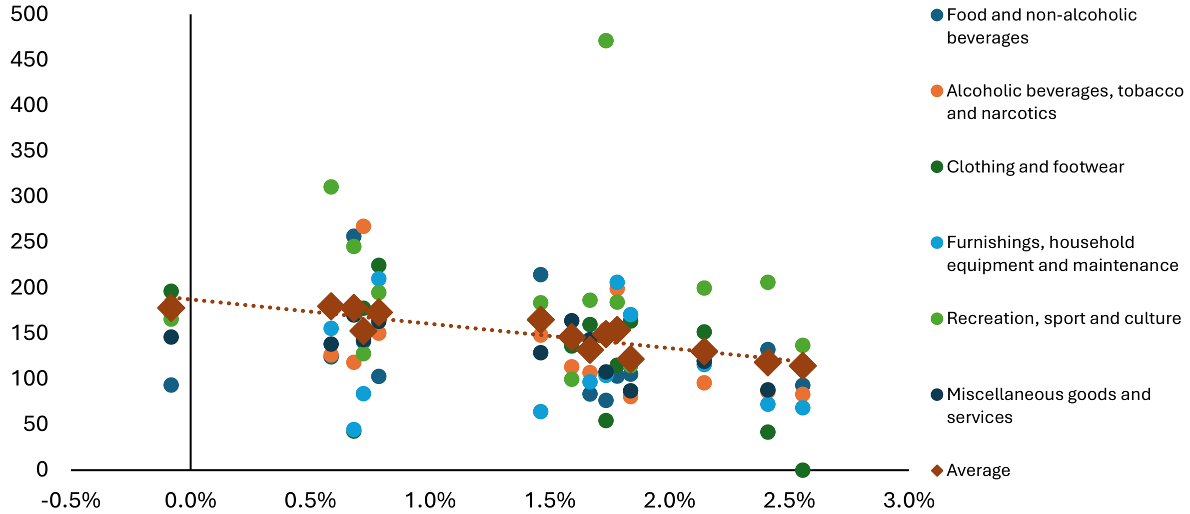
Sources: Supermarket data, BCRP.

When compared to quarterly inflation, duration correlates negatively (see Graph 14). The inflation regime could have a direct relationship with price formation, the update frequency, and the distribution of year-on-year percentage changes. Firms may perceive the need to update their prices more frequently due to cost increases or inflation itself, and even resulting in second-round effects. Therefore, prices become less sticky. It could be argued that this is related to a change from low to high inflation regime.

Quarterly inflation negatively correlates with duration of prices across all product division samples

Graph 14

A. Duration and quaterly inflation



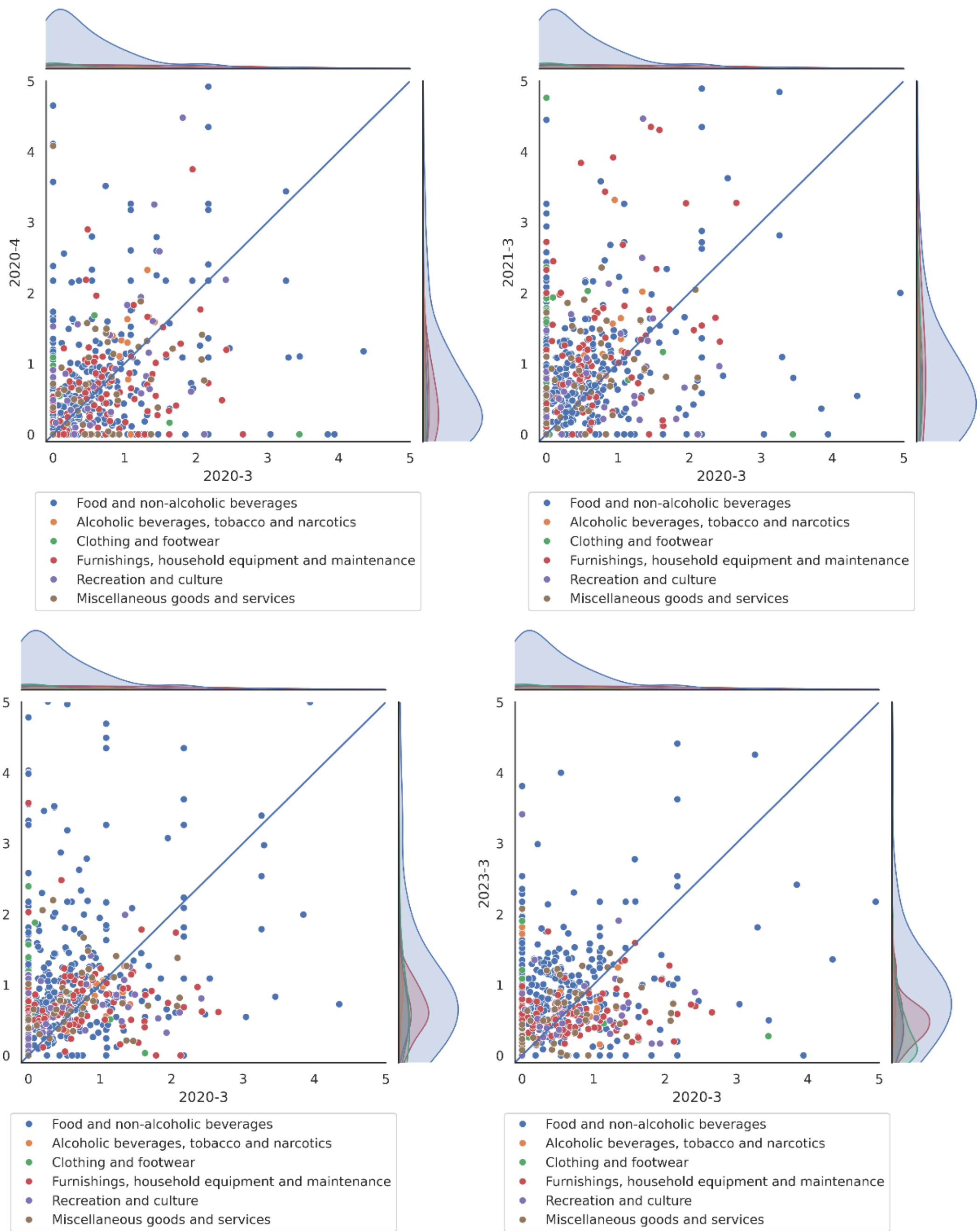
Sources: Supermarket data, BCRP.

Furthermore, at the product level, Graph 15 shows how the average quarterly update frequency increased from the third quarter of 2020. The 45-degree line and its vicinity indicate stability in this frequency. Comparing the third and fourth quarters of that year reveals a symmetrical dispersion, indicating variation in the degree of update without a dominant direction (increase or decrease). However, comparing it with data from the third quarter of 2021 shows an increase in observations above the 45-degree line, suggesting a higher proportion of categories with increased update frequency. In 2022, during the peak of inflation, categories above the line increased even more. With moderation in 2023, the pattern is closer to 2021 but still persistent. Additionally, it is crucial to note the persistence, indicating that even after returning to the inflation target range, the underlying behavior change could last longer.

During the inflationary period between 2021 and 2023, the change frequency increased compared to 2020

Graph 15

A. Change frequency by product category



Sources: Supermarket data, BCRP.

6. Conclusions

Online price data provides insights into their formation process and opens the door to constructing high-frequency indices. Moreover, it presents itself as a richer source of information, which, complemented with other data, allows for exploring how shocks propagate.

In the period 2020 to 2024, we observed different inflationary pressures, from the COVID-19 slowdown to the subsequent economic recovery. Indicators show that during this period, at the micro level, the proportion of products that changed their prices increased, moving the distribution of year-on-year price variations towards higher values until 2023, where it started to moderate. Also, price changes seem to be balanced between positive and negative values or biased towards the positive ones. This would imply that variations related to trend prices are bigger for increases than decreases, leading to inflation. Indicators show that price behavior varied as quarterly inflation grew. Price changes increased and that frequency has not returned to the 2020 distribution. This resulted in a negative correlation between inflation and duration of prices.

As part of the pending agenda, the classification of food and goods samples from supermarkets is being reviewed and improved to keep the quality of the data. Another important aspect is the generation of aggregate indices. This entails studying the series, analyzing differences when standardizing prices to the same unit or reviewing the correlation between products and inflation statistics of their corresponding category. With this, methodologies can be used to aggregate products at the sector level, and with the corresponding weights, aim to have a high-frequency CPI.

References

Aparicio, D., & Bertolotto, M. I. (2020). Forecasting inflation with online prices. *International Journal of Forecasting*.

Cavallo, A. (2013). Online and official price indexes: Measuring argentina's inflation. *Journal of Monetary Economics*, 60(2):152–165.

Cavallo, A. (2018). More amazon effects: online competition and pricing behaviors. *Technical report, National Bureau of Economic Research.*

Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*.

Coronado, H., Lahura, E., & Vega, M. (2020). Qué tan rígidos son los precios en línea: evidencia para Perú usando big data. *Technical report, Banco Central de Reserva del Perú*.

Gorodnichenko, Y., & Talavera, O. (2017). Price setting in online markets: Basic facts, international comparisons, and cross-border integration. *American Economic Review*, 107(1):249–282.

Joulin, A., Grave, E., Bojanowski, P., Nickel, M., & Mikolov, T. (2017). Fast linear model for knowledge graph embeddings. *arXiv preprint arXiv:1710.10881*.

Kiefer, N. M. (1988). Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, 26(2):646–679.

Lunnemann, P., & Wintr, L. (2006). Are internet prices sticky? *European Central Bank.*, Working Paper Series 645.

Macias, P., Stelmasiak, D., & Szafranek, K. (2022). Nowcasting food inflation with a massive amount of online prices. *International Journal of Forecasting*.

Nakamura, E., & Steinsson, J. (2008). Five facts about prices: A reevaluation of menu cost models. *Quarterly Journal of Economics*, 123(4):1415–1464.

Inflation and Price Dynamics Analysis using High-Frequency Data from Supermarkets

Evidence from Peru

Gonzalo Bueno Bustíos

Banco Central de Reserva del Perú

The opinions expressed in this study belong to the author and should not be attributed to the Banco Central de Reserva del Perú.

Contents

Inflation after 2020

- The inflationary process in Peru since the onset of the millennium has been characterized by a stable and low rate, akin to that of developed countries.
- Supply shocks were a major factor in the inflation surge in Peru observed since mid-2021, peaking in January 2023 (8.66 percent year-on-year).

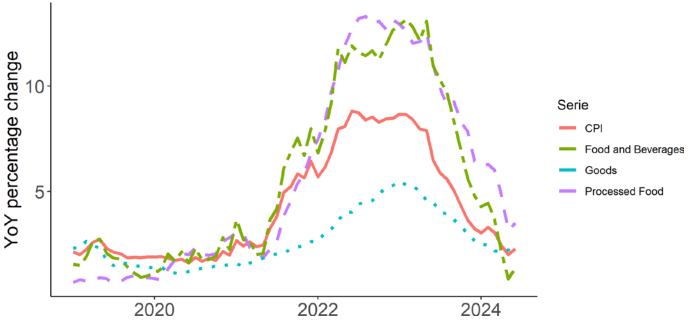


Figure: Year-on-year variations of price indices

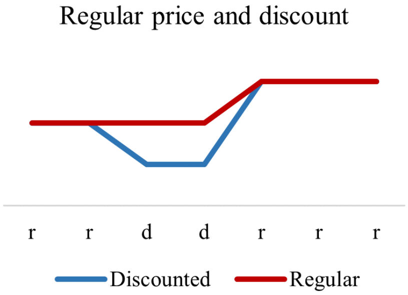
Data

Data was collected from supermarket webs using web-scraping:

- Items were divided between food, beverages and goods.
- They were classified according to the COICOP using machine learning techniques.
- The names of items had to be processed and simplified to optimize the machine learning algorithm.
- Around 134k products were considered in the sample of this investigation.
- A 30 day moving average was calculated to smooth price series.

Discount filter

- Extracted data includes regular and online (or sale) prices. The minimum between both is the effective price a consumer pays.
- The effective price still shows discount periods.
- To process the data, I developed a filter that removes sales and discounts. It is based on Nakamura & Steinsson (2008).
- If a price reduction lasts less than 30 days, then it is considered a discount and not a permanent change.

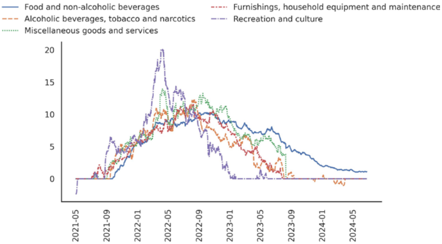


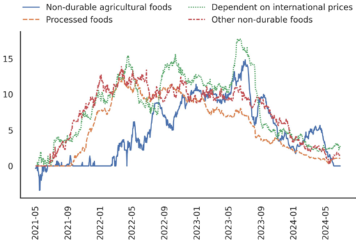
Contenidos

Prices increased similarly but decelerated asymmetrically

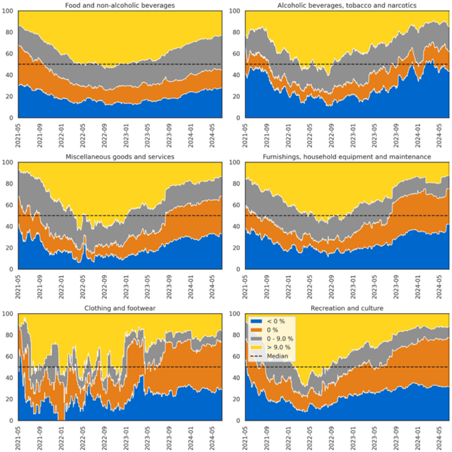
Year-on-year variation distribution

Median of year-on-year variations





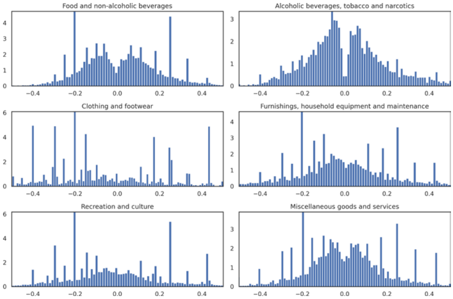
Year-on-year variation distribution

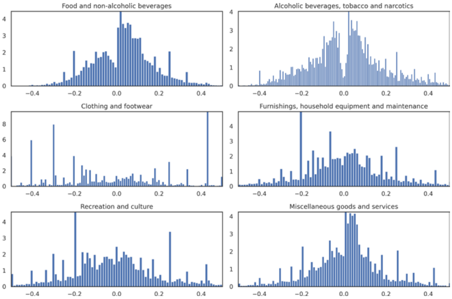


Persistent changes tend to be positive

Year-on-year percentage change magnitude

- Divisions seem to follow a bimodal distribution of changes. Small changes are less frequent.
- When removing discounts, as expected, positive changes have a higher presence.





Frequency of change

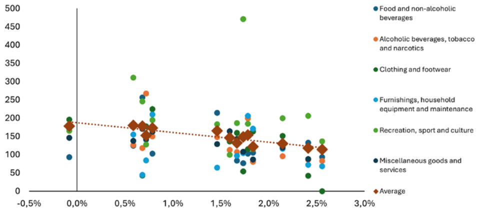
- Raw prices (without filtering) exhibit a rapid update.
- As the discount removal parameter increases, the implicit duration increases accordingly, therefore, prices seem to last for 7 to 13 months before moving upwards.

	Frequency of change (change probability)							
	Median				Mean			
Divisions	Regular	7	30	100	Regular	7	30	100
Food and non-alcoholic beverages	4.1%	2.3%	1.0%	0.7%	3.6%	1.9%	0.9%	0.6%
Alcoholic beverages, tobacco and narcotics	3.1%	1.9%	0.8%	0.4%	2.9%	1.9%	0.8%	0.4%
Clothing and footwear	2.3%	1.6%	0.6%	0.3%	2.5%	1.8%	0.6%	0.3%
Furnishings, household equipment and maintenance	2.3%	1.6%	0.7%	0.4%	2.4%	1.5%	0.7%	0.4%
Recreation, sport and culture	1.6%	1.2%	0.6%	0.4%	3.3%	2.1%	0.9%	0.5%
Miscellaneous goods and services	2.5%	1.7%	0.7%	0.4%	2.5%	1.7%	0.7%	0.4%

	Implicit duration (in days)							
	Median				Mean			
División	Regular	7	30	100	Regular	7	30	100
Food and non-alcoholic beverages	24	43	101	149	27	52	111	165
Alcoholic beverages, tobacco and narcotics	32	54	121	229	34	53	121	233
Clothing and footwear	43	62	155	392	39	54	157	399
Furnishings, household equipment and maintenance	43	62	139	263	40	65	142	235
Recreation, sport and culture	61	82	161	254	30	47	108	195
Miscellaneous goods and services	39	58	140	223	40	60	143	234

Duration changed with inflation

- There is a negative correlation between quarterly inflation and price duration.
- When shocks lead to persistent inflation, prices increase more and also adjust more quickly and frequently.



Frequency of change (median)

Division	Regular				7 days				30 days				100 days			
	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV
Food and non-alcoholic beverages	1.7%	4.0%	3.5%	4.0%	1.3%	2.9%	2.3%	2.3%	0.8%	1.3%	1.2%	1.0%	0.3%	0.9%	0.9%	0.5%
Alcoholic beverages, tobacco and narcotics	1.5%	2.1%	3.0%	2.8%	1.4%	1.7%	2.4%	1.4%	0.8%	0.6%	0.9%	0.6%	0.3%	0.3%	0.6%	0.2%
Clothing and footwear	0.0%	4.9%	1.8%	1.6%	0.0%	3.3%	1.2%	1.2%	0.0%	1.8%	0.7%	0.3%	0.0%	0.2%	0.3%	0.2%
Furnishings, household equipment and maintenance	1.7%	3.5%	3.1%	2.4%	1.2%	2.6%	2.1%	1.9%	0.6%	1.0%	1.0%	0.7%	0.3%	0.7%	0.6%	0.2%
Recreation, sport and culture	0.9%	1.2%	1.4%	1.7%	0.7%	1.0%	1.0%	1.3%	0.3%	0.2%	0.5%	0.6%	0.1%	0.1%	0.2%	0.2%
Miscellaneous goods and services	2.4%	2.3%	2.7%	2.9%	1.9%	1.9%	1.6%	2.0%	0.7%	0.9%	0.7%	0.7%	0.4%	0.7%	0.4%	0.4%

Implicit duration in days (median)

Division	Regular				7 days				30 days				100 days			
	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV	20-IV	21-IV	22-IV	23-IV
Food and non-alcoholic beverages	58	25	28	24	74	34	43	42	124	77	85	96	311	107	108	185
Alcoholic beverages, tobacco and narcotics	64	46	33	36	70	60	42	71	128	153	106	176	359	324	167	438
Clothing and footwear		20	54	61		29	80	80		54	138	388		636	384	582
Furnishings, household equipment and maintenance	58	28	32	41	84	39	48	53	155	104	104	147	298	136	154	494
Recreation, sport and culture	110	80	72	60	134	96	97	75	311	488	196	174	919	713	402	538
Miscellaneous goods and services	40	43	37	33	53	52	62	50	139	110	147	144	266	148	254	272

Conclusions

- **Price Formation Insights:** Online price data helps understanding price formation processes. It also enriches the analysis of how economic shocks propagate.
- **Price Changes:** The proportion of products with price changes increased, moving the distribution of year-on-year price variations toward higher values until 2023, after which it moderated. Changes were balanced or biased towards increases, contributing to inflation.
- **Behavioral Shifts:** As quarterly inflation grew, price changes became more frequent. This resulted in a negative correlation between inflation and price stability.
- **Data Quality and Index Generation:** Ongoing efforts include improving the classification of supermarket data and generating a high-frequency Consumer Price Index (CPI).

Inflation and Price Dynamics Analysis using High-Frequency Data from Supermarkets

Evidence from Peru

Gonzalo Bueno Bustíos

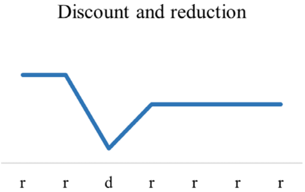
Banco Central de Reserva del Perú

The opinions expressed in this study belong to the author and should not be attributed to the Banco Central de Reserva del Perú.

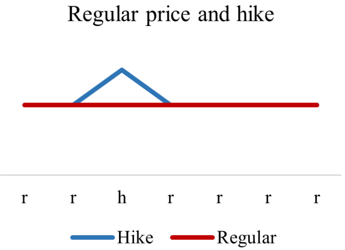
Appendix

Discount filter

Detection of v-shapes (30 days max. for discounts)

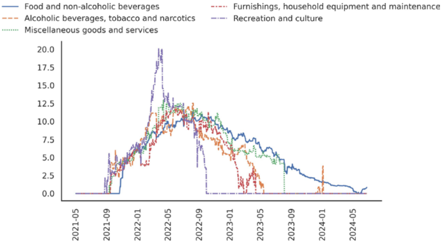


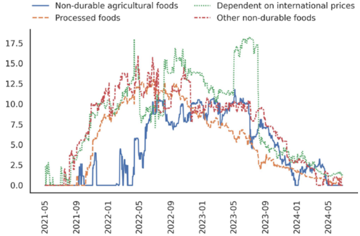
Price increase outlier (7 days)



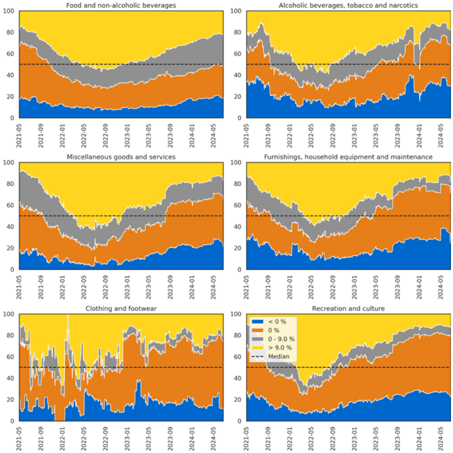
Variation distribution with filtered data

Median of year-on-year variations

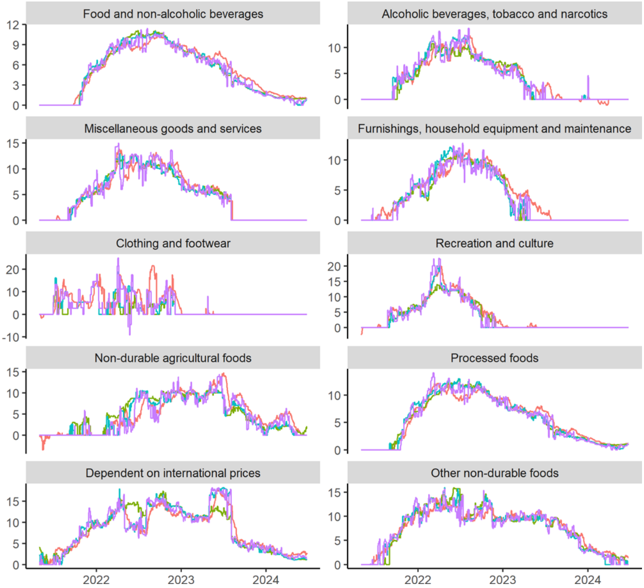




Year-on-year variation distribution



Discount filter parameter comparison



Frequency of change over different years