

Food Prices and Household Vulnerability in Italy: Insights from Web-Scraped Data

Ilaria Benedetti¹[0000-0003-3233-8543], Tiziana Laureti¹[0000-0003-1124-6949], Niccolò Salvini¹[0000-0002-3444-4094], Luigi Palumbo¹[0000-0002-8865-7116]

¹University of Tuscia, Italy, IT

Abstract. This paper investigates food price dynamics and inflation inequality in Italy by leveraging web-scraped data from major retailers. In the aftermath of the COVID-19 pandemic, surging food prices have disproportionately affected lower-income households, which allocate a greater share of their expenditure to essential goods. Drawing on over 5 million high-frequency price observations from 2021 to 2023, we examine within-category price variations across regions and time for staple food items, with a focus on Milk and Olive Oil. Using a modified version of the Time-Interaction-Region Product Dummy (TiRPD) model, we construct sub-national price indices across different price deciles, enabling a detailed assessment of inflation heterogeneity. Our results reveal that the lowest-cost items within each category experienced faster price growth compared to higher-priced counterparts—a dynamic known as “*cheapflation*”—thus limiting the ability of consumers to mitigate inflation by substituting down in quality. These patterns are consistent across product categories and regions. The findings underscore the need for timely and targeted policy interventions to mitigate the regressive effects of inflation and to protect vulnerable households.

Keywords: Inflation inequality, Cheapflation, Web scraping, Food prices, Household vulnerability

1 Introduction

The economic recessions and global inflationary surge following the COVID-19 pandemic have significantly strained household welfare, with widespread income losses exacerbating vulnerabilities, particularly among low-income groups (Midões and Seré, 2022). Beginning in the second half of 2021, inflation, primarily fueled by surging energy and food costs, rose sharply in Italy, peaking at 12.6 percent in October 2022 based on the Harmonized Index of Consumer Prices (HICP), its highest level since the mid-1980s. This surge in prices was partly offset by various government interventions, including broad-based measures for all consumers and more targeted support for lower-income households and individuals. Indeed, periods of high inflation determine strong redistribution patterns among different categories of consumers with a larger negative impact of inflation on purchasing power for poorer households (Curci et al., 2022).

Lower-income households, which allocate a disproportionately larger share of expenditure to essential goods (such as food), faced higher effective inflation rates than wealthier households (Corsello and Riggi, 2023).

Food expenditures accounted for an average of 18.4% of Italian household budgets in 2022, underscoring the sector's critical role in overall consumption patterns. By ranking households according to their equivalent expenditure and dividing them into five equally sized groups (quintiles), it is possible to observe significant differences across the expenditure distribution. Concerning food expenditure, less affluent households (first quintile) devoted approximately 25% of their total consumption to food, compared to just 14.4% among their more affluent counterparts (fifth quintile). During the same year, the HICP by expenditure quintile highlighted a pronounced inflation inequality, with low-spending households experiencing inflation rates 4.9 percentage points higher than their high-spending counterparts. This gap was primarily fueled by sharp increases in the prices of energy and essential goods, categories that weigh more heavily on constrained budgets. Although inflationary pressures began to ease in 2023, the continued presence of the expenditure gap underscores the uneven impact of inflation across households. More affluent families, who devote a smaller portion of their budgets to essential goods, have been relatively less affected by rising prices, highlighting how inflation has disproportionately burdened lower-income households (Corsello & Riggi, 2023). An immediate consequence of inflation heterogeneity is that price changes can exacerbate overall inequality not only through reduced real incomes, but also by unevenly affecting consumption patterns. It is therefore essential to understand inflationary dynamics, particularly for low-income households, which spend a larger proportion of their income on food and are more vulnerable to the adverse effects of rising food prices (Abay et al., 2023).

It is widely recognized in the inflation inequality literature that using highly disaggregated product-level price data is essential to fully capture the extent of inflation heterogeneity across different segments of the population (Jaravel, 2021). A growing body of research examines inflation inequality using high-frequency data. Cavallo (2020) and Jaravel and O'Connell (2020) provide real-time measurement of inflation inequality during the COVID-19 pandemic. A recent contribution by Cavallo and Kryvtsov (2024) highlights a phenomenon known as *cheapflation*, which refers to the disproportionately faster price increases of lower-priced goods within the same category. Their analysis of post-pandemic inflation dynamics shows that the internal price structure within product categories shifted significantly, limiting consumers' ability to alleviate inflationary pressure by simply switching to cheaper options. Similarly, Chen et al, (2024) used detailed UK household purchase data to show that between 2021 and 2023, lower-priced products experienced significantly higher inflation than more expensive ones, exacerbating inflation inequality across households.

With the aim of exploring this phenomenon in Italy, we use web-scraped data to analyse food price dynamics and inflation inequality. Our analyses rely on highly disaggregated data at the product level, offering valuable insights into inflation heterogeneity within narrowly defined goods categories. This approach allows a detailed examination of the price distribution within the same product category across

different percentiles of the price distribution (e.g. less expensive versus most expensive brands), such as flour, milk, olive oil, pasta and rice, and to compare their relative prices over space and time.

By leveraging web-scraped data from major Italian retailers—comprising over 5 million price observations across staple categories (like flour, milk, olive oil, pasta, rice) we use a modified version of the Time-Interaction-Region Product Dummy (TiRPD) model (Aizcorbe & Aten, 2004), to analyze price dynamics across quality tiers (proxied by price deciles) to quantify the unequal burden of inflation. This model allows the simultaneous modelling of time and space and is therefore able to fully capture price dynamics. This research provides a pioneering spatio-temporal analysis by focusing on specific price aggregates for each product category: the lowest decile and highest price decile for each product identified within each outlet. The results of this research can inform policy decisions aimed at mitigating the negative impact of inflation on vulnerable populations from adverse effects of inflation, ensuring that interventions are timely and targeted.

2 Data and Methodology

Our web-scraped dataset spans from July 2021 to February 2023, focusing on 167 outlets in 7 different regions affiliated with a major Italian retail chain, that is Calabria, Campania, Emilia-Romagna, Lazio, Lombardy, Sicily and Umbria. We developed web scraping routines to perform a daily collection of prices and details for products available in each individual outlet, leveraging a hidden Application Programming Interface (API) in the retail chain e-commerce platform. The dataset includes about 1 million 350 thousand observations. Although the analysis encompasses multiple product categories, this paper presents results for two representative groups—Milk and Olive Oil—with individual items identified through Global Trade Item Numbers (GTINs) which are the unique identifiers used to distinguish products globally, making it possible to track and identify specific items regardless of location or retailer.

We only selected products with sizes equal to 1 liter, in order to keep a direct comparability of prices amongst our sample. The pricing concept refers to the monthly price for each GTIN in each outlet. Our data collection amounted to 521 unique products for Milk and 102 unique products for Olive Oil, with a total of 189,445 and 82,405 monthly observations, respectively. Table 1 shows the characteristics for the selected product categories—milk and olive oil—providing insights into their market representation and share of total food expenditure.

Table 1. Descriptive statistics for product categories

Category	Product size	# products	# observations	% of food expenditure
Milk	1 l	521	189,445	1.96
Olive oil	1 l	102	82,405	0.98

To illustrate the evolution of price distribution over time, Figure 1 reports results for the Milk product group as a representative example. The figure shows notable shifts in price levels, particularly within the lowest price deciles, throughout the study period. Although only the Milk case is presented here for illustrative purposes, similar patterns—namely, upward shifts especially in the lower decile—are observed across all product categories included in the analysis.

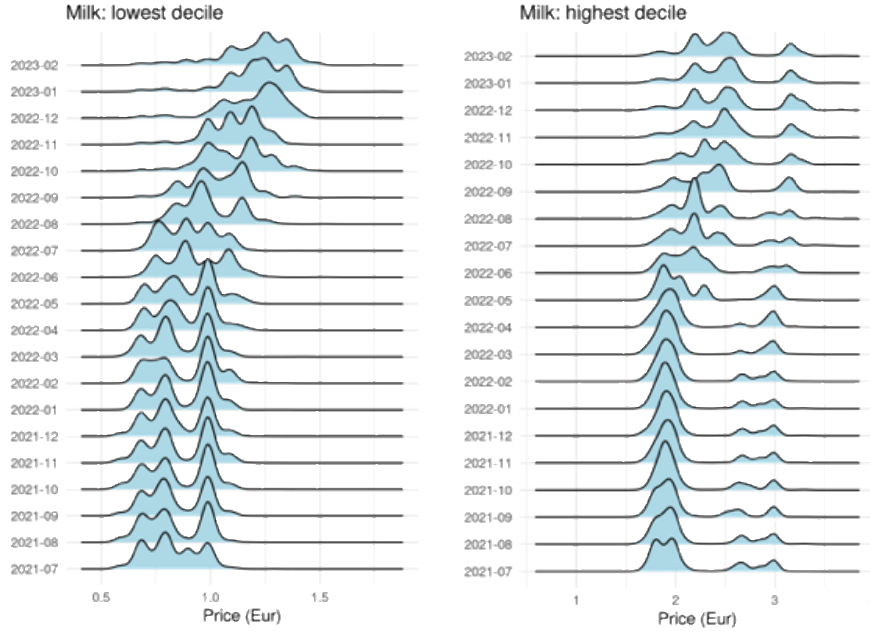


Fig. 1. Price distribution for Milk product category over the period 2021-2023

To explore the price distribution within the same product category across space and time by distinguishing least-cost food products price evolution, we applied a modified version of the TiRPD index, introduced by Aizcorbe and Aten (2004) in the context of the stochastic approach to price indexes considering a single homogeneous product across all outlets:

$$\ln P_{ijt} = \sum_{t=1}^T \sum_{j=1}^J \delta_{jt} C_{ij} T_{it} + v_{ijt}$$

Where $\ln P_{ijt}$ and D_{ijt} are respectively the log-price and the dummy for the product in outlet i in region j at time t ($i = 1, 2, \dots, N$; $j = 1, 2, \dots, M$; $t = 1, 2, \dots, T$). C_{ij} and T_{it} are dummy variables for each combination of area and time. The sub-national price index for the area j at time t is given by $\exp(\delta_{jt})$. In our case, all products in the same price class – least-cost and highest – are considered homogeneous.

3 Results

Results from TiRPD models are reported in Figure 2 and Figure 3. Findings from least-cost food (lowest decile) products analysis revealed that, on average for these two food product groups, the prices of the least expensive items have increased at a faster pace compared to the prices of the most expensive products in the same category, both over time and across different regions in Italy.

From Fig. 2 it is clear that the prices of milk products in the lowest cost distribution have increased more than those in the highest decile even if the trend is relatively homogeneous across the regions analysed. This suggests that, despite a general increase in prices, the trend has been fairly uniform across the regions studied, indicating a common pattern in the behaviour of dairy product prices.

For the flour product group, the most significant price increase is observed in the lowest-cost product group (lowest decile), particularly in the region of Sicily, where the price index reaches its peak (with TRPI equal to 168) in August 2022, compared to the base period of July 2022 in Lazio.

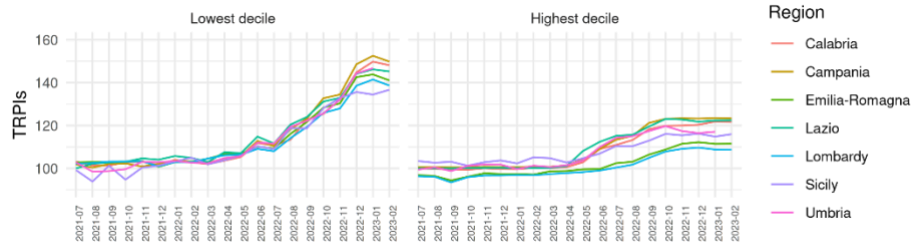


Fig. 2. Spatial and Temporal price index for Milk product group according to price band

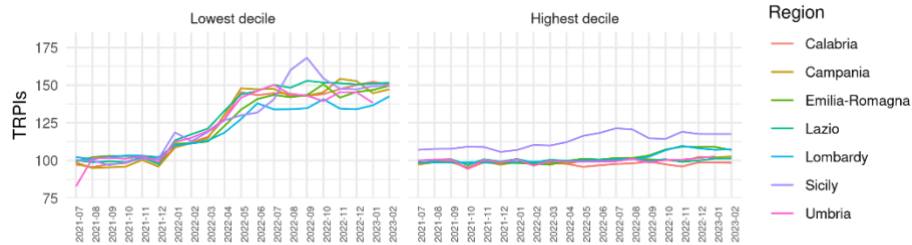


Fig. 3. Spatial and Temporal price index for Flour product group according to price band

Overall, the notable upward trend observed in the lowest cost distributions, should be related to the fact that these products often rely on basic raw materials and offer lower profit margins. Therefore, retailers may have less flexibility to absorb cost increases, resulting in more pronounced price growth. Cavallo and Kryvtsov (2024) pointed out other reasons, including targeted fiscal stimulus, which likely increased the demand for cheaper varieties, and the possibility that cheaper products tend to depend more on global supply chains, like the ones disrupted by COVID-19. The phenomenon of *cheapflation* has particularly adverse effects on vulnerable households, as these groups depend more heavily on lower-cost goods—precisely the items experiencing the

most pronounced price increases. Given that such households already devote a substantial share of their income to essential goods, their ability to adjust consumption patterns or substitute toward cheaper alternatives is limited. Consequently, *cheapflation* leads to a more significant erosion of real purchasing power for these groups, thereby intensifying existing socio-economic disparities and amplifying the regressive nature of recent inflationary trends.

References

1. Aizcorbe, A., Aten, B.: An Approach to Pooled Time and Space Comparisons. SSHRC Conference on Index Number Theory and the Measurement of Prices and Productivity, Vancouver, Canada, pages 1–19 (2004)
2. Cavallo, A., Kryvtsov, O.: Price discounts and cheapflation during the post-pandemic inflation surge. *Journal of Monetary Economics*, 148, 103644 (2024).
3. Chen T., Levell P., O'Connell, M.: Cheapflation and the rise of inflation inequality. 36. London: Institute for Fiscal Studies. <https://ifs.org.uk/publications/cheapflation-and-rise-inflation-inequality>, last accessed 2025.04.11
4. Corsello, F., Riggi, M. Inflation is not equal for all: the heterogeneous effects of energy shocks. Banca d'Italia Eurosystem (2023)
5. Curci, N., Savegnago, M., Zevi, G., Zizza, R.: The redistributive effects of inflation: a microsimulation analysis for Italy. *Bank of Italy Occasional Paper*, 38, (2023).
6. Drenik, A. Perez, D. J.: Price setting under uncertainty about inflation. *Journal of Monetary Economics*, 116:23–38 (2020)
7. Jaimovich, N., Rebelo, S., Wong, A.: Trading down and the business cycle. *Journal of Monetary Economics*, 102:96–121 (2019).
8. Jaravel, X.: Inflation inequality: Measurement, causes, and policy implications. *Annual Review of Economics*, 13(1), 599–629 (2021).
9. Jaravel, X., O'Connell, M.: Real-time price indices: Inflation spike and falling product variety during the Great Lockdown. *Journal of Public Economics*, 191, 104270 (2020).
10. Kaplan, G. and Menzio, G.: The morphology of price dispersion. *International Economic Review*, 56(4):1165–1206 (2015).
11. Midões, C. and Seré, M.: Living with reduced income: an analysis of household financial vulnerability under covid-19. *Social Indicators Research*, 161(1):125–149 (2022)
12. Sheremirov, V.: Price dispersion and inflation: New facts and theoretical implications. *Journal of Monetary Economics*, 114:59–70 (2020)