Economics Observatory.

Economic and Social Research Council

Faster, bigger, cheaper:
how AI can improve UK price data

EMBARGO –ESCoE CONFERENCE – THURSDAY 16TH MAY

Richard Davies, Finn McEvoy
Economics Observatory

May 2024
www.economicsobservatory.com

Faster, bigger, cheaper: how AI can improve UK price data

Richard Davies† and Finn McEvoy‡

† London School of Economics, University of Bristol, Economics Observatory

‡ LSE and Economics Observatory

§ Working Paper – 16th May 2024.

This paper introduces a new dataset of 25 million daily prices collected from seven major British supermarkets which we use to develop a novel index of grocery price inflation. We access data from the websites of retailers representing over 80% of UK grocery expenditure. The prices offer insights into the frequency, timing, and distribution of price changes, providing new evidence on pricing behaviour and rigidities, core elements of macroeconomic models. Compared with existing UK microdata, our dataset is larger, contains a far wider set of variables (including, for example calorific content and user ratings), and is more timely, available at a lag of around 12 hours. The scale and breadth of the data presents a new problem: product classification. Here we use a novel Large Language Model-driven (LLM) methodology, linking items to their respective COICOP codes and items in the Consumer Prices Index (CPI) basket. We are able to investigate price rigidities and inflation with unprecedented granularity, producing daily estimates. The results, and the dataset, have a wide range of policy uses.

∗ Economists at HM Treasury, the Bank of England and the Office for National Statistics provided helpful comments that have helped us shape this project. Participants at the LSE's POID seminar series also provided valuable comments. We are grateful in particular for ideas from Adam Bricknell, Tara Murphy, Ahmet Aydin, Liam Greenhough, Elishama Tizora, Bradley Speigner and Peter Lambert. Denes Csala and Josh Hellings provided help and guidance with the data engineering aspect of the paper. James Kean and Hannah Cantekin provided excellent research assistance. All errors, omissions and opinions in the paper are our own.

Contents

Executive summary

1. Britain is emerging from a bout of inflation the severity of which has not been seen since the 1970s. This has raised vital questions about the sources of inflation and its likely duration, or stickiness. It is also a reminder that longstanding puzzles in macroeconomics—the frequency and size of price changes, and the role of menu costs—evolve. The latest period of high inflation also raises interesting questions about the technology available to policymakers: could early warning systems be put in place to alert policymakers to emerging price pressures?

2. This paper introduces a new and growing dataset of daily grocery prices to help answer these questions. We collected 27 million prices between July 2023 and April 2024 from seven supermarkets. The retailers account for around 80% of UK groceries sales (Kantar, 2024). The prices collected cover food and drinks, alcohol and tobacco; and cover between 7,000 and 30,000 product lines at the supermarkets. We use the prices to examine the sources of inflation, to study price rigidity at the firm level, and to build a new automated index of grocery price inflation which is available daily. We make four distinct contributions.

3. **Price collection**. First, we set out a new algorithm for the collection of grocery prices. We outline three approaches to automated price collection, explaining their strengths and weaknesses and the efficiency gains from our preferred approach. Our automated and distributed program delivers around 100,000 prices per day. The data is far richer than previously available, including consumer prices along with unit prices, loyalty discounts, user ratings, alternate categories, regional availability, product limits, the page location of the item, and nutritional information. The approach is also fast: prices are collected and available for analysis with a lag of around 6 hours.

4. **Price classification using LLMs**. With our new dataset in hand, we are faced with a daunting task: classification. Due to the huge ranges of products that UK supermarkets stock, we have over 150,000 shop-product pairs to parse and categorise so that other data—for example, expenditure weights or trade exposures—may be merged in. Our second contribution is a method for parsing the data using LLMs. We adopt a three-stage filter using the GPT-4 model produced by Open AI. This approach will be set out in detail in a background paper aimed at statisticians and developers (Davies, Hellings and McEvoy, forthcoming).

5. **Price changes: new facts on frequency, size, and timing.** Our cleaned and classified price data allows us to establish a range of new facts on the frequency, size, and timing of

price changes. These questions are core to the operation of macroeconomic models and until now have been assessed at a monthly frequency in the UK. We find:

- Frequency. Price change frequencies depend on the item in question. Most goods change price rarely. On average 7.7% of prices change each month, with a mean of 56 days between changes. Some items change price much more regularly: over six months some items have changed price 18 times.

- Sizes. Most retailers change prices according to a bimodal distribution with few small tweaks. Sales appear to play a strong role in price changes: 20% price cuts and 25% price rises are the most common changes.

- Timing. There is heterogeneity in the timing of price changes by firms. Some retailers change many prices every day. For others, price changes are highly synchronised, with most price alterations occurring on a just a few days. The latter pattern is consistent with menu costs that create an economy of scope in price changes.

6. An automated food CPI. Our final contribution is to produce an automated food CPI for the UK. The approach has been used in a number of settings with automated estimates reaffirming (Poland, Macias et al., 2023) and challenging (Argentina, Cavallo, 2012) official figures. Our findings contribute to the evolution of UK price measurement. We are able to track the UK CPI basket by adopting a set of tight matching criteria. This 'narrow' measure may be useful for policymakers in periods when prices are moving quickly, or when it is unclear the extent to which price shocks may be passed through. We also create a 'broad' measure which matches the CPI basket and allows for substitutes. The broad measure diverges from the current inflation measure, showing how the backet choices can influence inflation statistics. Both measures should be seen as complements, rather than substitutes, to the official CPI; the Auto-CPI could potentially be an early indicator used alongside the ONS official measure.

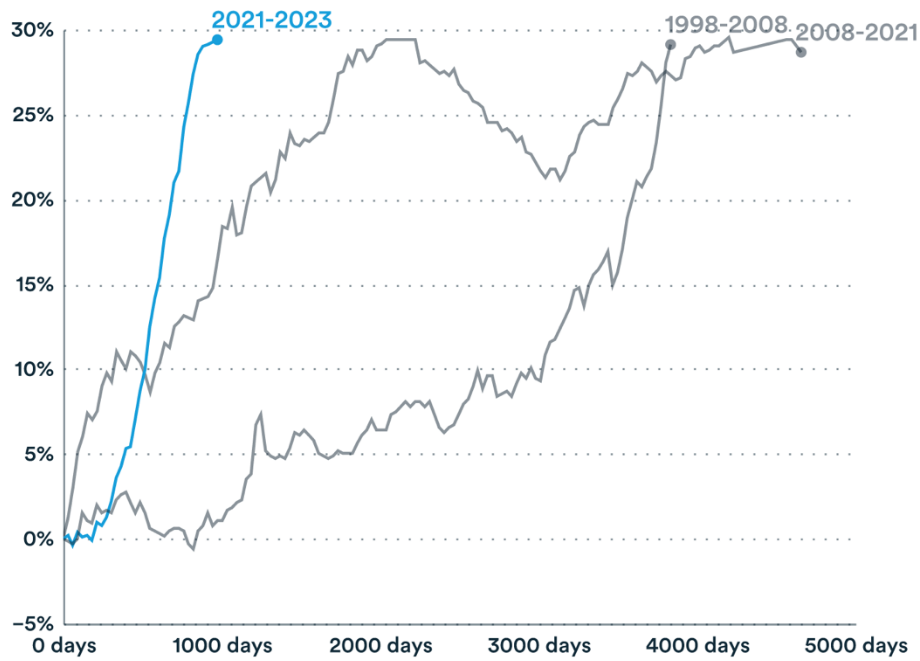
7. The paper proceeds as follows. Section 1 reviews the state of play in the UK and reviews other research using price microdata. Section 2 explains the automated collection of 100,000 prices per day. Section 3 sets out our algorithm for categorisation and classification of supermarket products using an LLM. Section 4 presents new results on price flexibility in the UK, and our automated CPI. Section 5 presents conclusions and next steps, including setting out how policymakers and researchers might use our data.

I Introduction and literature

8. Food prices increased by 30% between January 2021 and October 2023—a period of 33 months (Figure 1). The previous rise of the same size took 13 years (2008 to 2012). This extraordinary increase in inflation has led to many research and policy questions about the sources of price rises in the UK (Bank Underground, 2024).

Figure 1: 13 years of food price inflation in 33 months

Cumulative increase in the price of food and soft drinks



Notes: The 2021-2023 series ends after 1003 days in October 2023. Source: ONS, authors' calculations

9. Following an economic shock price adjustment can be fast. In the UK, for example, VAT changes, the Global Financial Crisis, and pandemic-related policies have all led to sudden bursts of re-pricing (Davies, 2021). Price collection, however, relies on brick and-mortar stores, meaning that such price shifts only appear in official micro data with a lag of 30 days or more. Following shocks, whether policy-led or economic, decision-makers and analysts can therefore end up working with outdated information.

10. Brick-and-mortar price collection leads to understandable and unavoidable lags. To produce the CPI, the ONS uses grocery prices collected during a few days in the middle of each month which are then incorporated into inflation figures released towards the end of the following month. This has two major implications for the scope and timeliness of price statistics. Firstly, a short price collection window means that price collection incorporates

prices at a single point of time, rather than capturing data representative of the whole month – the aim of inflation statistics (IMF, 2020, 5.42). Secondly, the lag between collection and reporting results in figures that are non-current, with inflation data reflecting the situation from over a month prior.

Figure 2: CPI Data Collection and Reporting

The lag between price collection and reporting

Notes: The ONS targets price collection on a single 'index day' around the 2nd or 3rd Tuesday of each month.

Source: Authors' calculations.

11. During high-inflation periods, this lag means decision-makers and analysts are working with outdated information. As others have done outside of the UK, we can estimate inflation using data that is just days, not months, old. Modern data collection and parsing methods—automated HTTP requests and LLMs—can help us do this. These tools also bring other benefits. The first is cost: while our methodology does rely on the provision of cloud computing, it is likely to be considerably cheaper than conventional methods. Another is breadth and detail: for some of our sources, we are able to access far richer data on UK goods by collecting user reviews, calorific content, suggested substitutions and regional availability along with prices. All of these have economic applications. This paper sets out the first results from a new project which aims to provide a fast and cheap source of UK pricing data while also producing bigger (and richer) datasets.

The UK experience with price scraping

12. The ONS has run several web scraping initiatives to enhance its price statistics. Beginning with a 2014 pilot (part of the Big Data Project), it aimed to generate price statistics for 40 CPI items using a commercial partner's back series and additional scraped data, as documented by Naylor et al. (2014). Progressing to a 2015 project, the ONS refined its methodology by classifying prices based on substring matching, though this approach faced several classification and technical hurdles, noted by Breton et al. (2015). Subsequently, Metcalfe et al. (2016) introduced the CLIP methodology, which approached price changes through clusters of products based on their information and pricing, although this project is no longer active. In a now concluded pilot, running from April 2021 to September 2022, the ONS monitored the prices of 30 inexpensive items using web-scraped data and observed minimal inflationary discrepancy between the least expensive items and their standard

counterparts. Most recently, experimental work has been carried out to scrape and classify online clothing prices (ONS, 2020).

13. The ONS is not currently using web-scraped pricing data in official statistics, but it has begun incorporating other digital pricing microdata, as part of its 'Transformation of consumer price statistics' plan. Since 2023, transaction-level data from the LENNON system has been utilised for rail fares, providing comprehensive details on the cost and quantity of tickets sold across Great Britain. In 2024, the ONS has plans to integrate alternative big data sources into other areas, beginning with second-hand cars and private rents (ONS, 2023).[1]

Price scraping overseas

14. The international experience with price microdata is set out in Davies (2021), with useful updates in Bank of England (2024). Central banks and statistical agencies have also made progress using automated price gathering, including scraping.

15. An early pioneer of the use of price microdata for inflation estimation was the MIT/Harvard Billion Prices Project led by Alberto Cavallo and Roberto Rigobon. Across 50 countries, the project involved the collection of more than 5 million prices per day from 300 retailers (Cavallo and Rigobon, 2016). Using web-scraped data, the authors were able to pre-empt movements in official figures and provide evidence of manipulated statistics. For instance, in the United States, nowcasted inflation estimates were able to pick up a 1% fall prices following the October 2008 collapse of Lehman Brothers (ibid.). And in Argentina, their online price index indicated inflation three times as large as official figures, confirming claims of manipulation (Cavallo, 2013). The project has since concluded but is succeeded by the private research firm PriceStats.

16. In Europe, the European Central Bank's PRISMA[2] Network has collected and spurred macroeconomic microdata research. Established in 2018, the project has brought together projects involving web-scraped prices, scanner data, and other microdata. This collaboration has sparked a series of web-scraping and scanner data projects on inflation in several countries, including Germany (Menz and Wieland, forthcoming), Austria (Messner and Rumler, forthcoming; Beer et al., 2023) and Denmark (Dedola et al., 2019), with Porqueddu et al. (forthcoming) extending the research across Europe.

17. Within the PRISMA network, Poland's central bank, Narodowy Bank Polski, has led the most extensive academic research efforts. From 2009 to 2020, Macias et al. (2023)

[1] We would like to thank Liam Greenhough for his helpful explanation of the ONS' data transformation.
[2] Price-setting Microdata Analysis Network

gathered 650 million prices from eight Polish grocery stores to develop automated CPI estimates for Poland. By integrating machine learning classification with significant human correction, they attained a 0.77 correlation with official statistics. Additionally, this dataset has supported further studies, such Stelmasiak et al. (2023), which examined product availability during COVID-19 and found that reductions in availability led to only a modest increase in prices. With a similarly structured market and grocery focus, this project provides a strong benchmark for British efforts.

18. Other recent projects have found some predictive power in novel and alternative data sources. Nakajima et al. (2021) analyse sentiment from around 2,000 individuals in the Japanese Cabinet Office's Economy Watchers Survey to predict inflation trends. Similarly, Eugster and Uhl (2024) examine the sentiment of 700,000 American news articles over nearly 20 years. Although these methods provide leading indicators of inflation, they also introduce a notable amount of noise into the predictions.

19. Moving beyond academic projects, some statistical bureaus have begun integrating web-scraped prices into their official inflation figures. The Australian Bureau of Statistics, for instance, collects approximately 1 million prices weekly from 65 retailers, focusing on items that are either easily classifiable, like car parts, or difficult to collect manually, such as clothing. As of 2020, about 5% of Australia's price data is sourced from web-scraped information and another 16% comes from scanner data (ABS, 2020), with the move towards automation faciliatating more frequent CPI releases. Similarly, from 2021, Austria has supplemented its manual price collection efforts by incorporating web-scraped prices (Statistics Austria, 2023). This new method was first applied to rent data and later expanded to cover electricity and gas, municipal fees, mobile phone tariffs, and clothing. Starting around 2013, the United States has taken tentative steps in this direction, with the Bureau of Labor Statistics (BLS) first employing web-scraped data for specific Medicare drug statistics (Horrigan, 2013).

II Collecting prices

20. The data we collect are daily records of UK grocery prices from British supermarkets. To enable the collection of daily observations, prices are collected programmatically from the retailers' websites through a process commonly known as scraping. Since July 2023, we have collected over 25 million prices and now measure 95,000-105,000 items per day (Figure 3). These prices are sourced from seven of the UK's ten largest supermarkets: Tesco, Sainsbury's, ASDA, Morrisons, ALDI, Waitrose, and Iceland.

Table 1: Three approaches to price collection				
Technique	*Speed*	*Cost*	*Availability*	*Dimensionality*
HTML Scraping	Fast	Low	Low (1/7)	Low
Screen Scraping	Slow	High	High (7/7)	Low
AJAX Requests	Fastest	Low	Medium (4/7)	High
Source: Authors' calculations, adapted from Macias et al. (2013).				

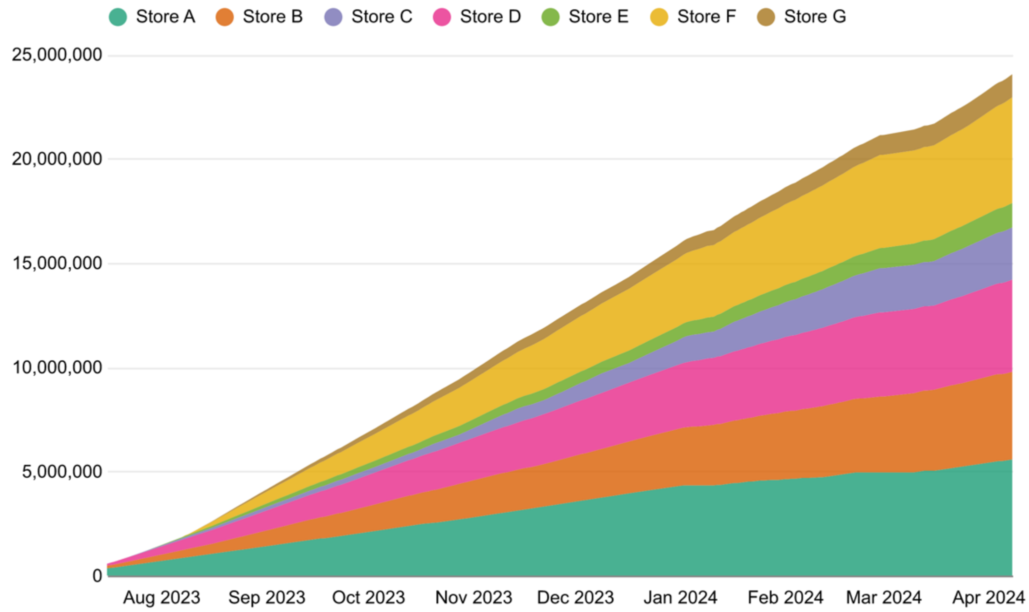
21. As the websites differ in their structure, so too must the technical 'stack' (the set of technologies) used to extract prices. Three separate approaches are used to extract prices with multiple implementations for some to build in redundancy. These three techniques – scraping from page source, screen scraping, and intercepting network requests – are summarised in the table above, adapted from Macias et al. (2023).

22. The first approach involves making HTTP requests to product listing pages and extracting product data directly from the returned HTML. Typically, a single request will be made which returns HTML intended to be rendered by a web browser. Using packages such as BeautifulSoup, the page source is parsed, and product information is extracted by looking for known page elements. This method allows for efficient extraction of data without the need to render the visual interface, making it less resource-intensive and allowing the scraping to be performed quickly using low-rent servers.

23. However, this conventional approach only works for few websites. In particular, it allows only the parsing of static websites where product information is returned in the initial page response. This is typically not the case. In their initial responses, most modern sites return a basic page layout along with code (JavaScript) that then loads product information, after a short delay. For example, some retailers use an 'infinite scroll' interface where products are requested in small batches when a user has scrolled to near the bottom of the page. This system in represented in in the additional figures. Having received the request, the server passes back the requested product information in its response which the page's code then uses to create HTML elements displaying the new information.

Figure 3: Prices collected per day

July 2023 - April 2024



Notes: Where multiple price observations are recorded for a single product in one day, only one observation is counted. Anonymised store names are not constant across figures. Source: Authors' calculations.

24. To collect prices from sites that load prices in this way, tools such as Selenium which allow programmatic control of web browsers can be used.[3] The supermarket sites are loaded to a browser which receives commands to navigate through the page. For example, in the case of scraping prices from an 'infinite scroll' site, instructions are sent to periodically scroll down until the last product is loaded. Screen scraping is possible for virtually all retailers.

25. The final method—request scraping—is a more direct version of the second. By interacting with retailers' websites, the endpoints and structures of data requests are revealed. Once these are known they can be made without a browser. Since data requests have standard and (usually) comprehensible form, they can be intercepted and repeated. Of the three methods, request scraping is the fastest and least computationally intensive.

26. The methods also differ in the depth of data returned. Retrieving product information from the HTML typically returns just the information displayed to the user on the page: names, prices and a url (from which the product id can be extracted). However, more dimensions are usually contained in the AJAX responses collected in request scraping. This

[3] Source code available at: https://github.com/SeleniumHQ/selenium

extra information (which is typically included to allow allowing customers to sort or filter by brands, subcategories or ratings) provides a richer picture of the products. These extra dimensions vary by supermarket but include alternate categories, ratings, regional availability, product limits and nutritional information.

27. Since we are collecting prices daily, our dataset evolves constantly. At the end of April 2024 our dataset looked as set out in Table 2 below.

Table 2: Summary of dataset

Supermarket	*Prices (m)*	*Products*	*Days*
Store A	5.9	37,253	256
Store B	4.5	23,356	284
Store C	5.5	43,290	244
Store D	4.6	18,705	270
Store E	1.2	9,185	259
Store F	3.9	35,172	221
Store G	1.2	8,670	224
Total	**27.0**	175,631	-

Source: Authors' calculations, Kantar (2024).

III Classification

28. The dataset of supermarket products has huge potential to reveal price trends. For example, insight can be gained simply by looking at aggregate movements. As Davies (2021) shows, the balance between price cuts and rises alone is a simple and strong predictor of CPI. However, creating an inflation estimate requires us to weight products according to their importance in consumer expenditure. These weights are produced annually using the Household Final Consumption Expenditure survey (HHFCE). The ONS releases COICOP subclass and item weights but store identities and weights are anonymised. Instead, we use Kantar (2024) market-share estimates.

29. We also want to use the new scraped microdata as a complement to existing CPI data. For these reasons we need to match our products (e.g., 'Tesco Soft White Rolls 6

Pack') to both subclasses (e.g., 'Bread') and CPI items (e.g., 'SIX BREAD ROLLS-WHITE/BROWN').

30. Both in the UK and internationally, products are indexed according to the UN Statistics Division's COICOP (Classification of Individual Consumption by Purpose) system (United Nations Statistics Division, 2018). At its core, the COICOP system divides consumption items into several levels: divisions, groups, and classes. Each level offers a more detailed and specific categorisation, subdividing its parent. In the United Kingdom and other European countries, this system is further subdivided according to Eurostat's ECOICOP (Eurostat, 2016), also known as COICOP5 or COICOP subclass. Finally, below the ECOICOP categories are ONS-specific items (Table 3).

Table 3: Product classification

Level	Name	Example
1	Division	Food and non-alcoholic beverage (01)
2	Group	Food (01.1)
3	Class	Bread and Cereals (01.1.1)
4	Subclass	Bread (01.1.1.3)
5	Item	SIX BREAD ROLLS-WHITE/BROWN

31. We classify our supermarket products at the COICOP subclass and Item level with a three-step process:

- **Product Categorisation with an LLM**. First, we categorise products into COICOP subclasses (e.g., 'Bread'). To do this, we first categorise products to an intermediate system which then maps to COICOP codes. These categorisation choices are made, and verified, with many LLM queries. This yields a category system with both CPI items and supermarket products.

- **Generating Item candidate shortlists with embeddings**. Second, for each CPI item, we generate a list of candidate supermarket products. We cast a wide net, including products that may be imperfect matches. To do this, we use embeddings – vectors representing the semantic meaning of data – to find the products that are most similar to each CPI item.

- **Selecting the best candidates**. Finally, we select only the best matches from our shortlist of candidates. Using carefully constructed LLM-prompts, we assess each candidate product and keep only the best fits.

Stage 1 – LLM categorisation and verification

32. First, we map supermarkets' products existing within that firm's category system to an intermediate category system. This intermediate system is then used to map to COICOP classes. Retailers assign their products to category systems, which differ by store. Some subdivide many times into lots of small groups (across all layers of its hierarchy, one store has almost 2000 categories) while others' systems are far sparser. For many, the division of products reflects their physical groupings in stores: products tend to be initially categorised by the aisles in which they would appear in-store (e.g., 'Fresh', 'Frozen', 'Bakery', 'Store Cupboard') then divided further by product types.

33. The COICOP system has a broader scope and divides products differently. In this system groceries are more likely to be categorised by their food content (meat, vegetables, cereals) than their role (pie, burger, sauce) or by their packaging/preservation (fresh, frozen, tinned). This leads to divisions that can seem unintuitive:

34. Similar items exist in disperse codes. For example, pies including meat are classed in 01.1.2.8 ('other meat preparations'), while quiches are found in 01.1.1.5, alongside pizzas.

35. Single codes include disperse items. Items that seem fundamentally different exist under the same codes: The category 01.1.7.3 ('dried vegetables') contains both vegetable burgers and canned tomatoes.

36. This presents a challenge for unifying the CPI items and supermarket products. Re-categorising products into a common system is easier (and better performed by LLMs under the process explained below) than re-categorising into one that is structured dissimilarly. Therefore, to aid in categorisation, we have constructed an intermediate, COICOP-like system that mirrors the category systems of supermarkets but maps to COICOP codes. As with the supermarket systems it builds upon, our intermediary initially divides by the aisle-like groupings found in supermarkets ('Fresh Food', 'Food Cupboard', 'Frozen', etc.), but carefully subdivides into categories which efficiently segregate items according to their placement in the COICOP. At the end of our process, supermarket items are either mapped to an COICOP code or excluded.

37. To place products into this category system, we use a Large Language Model (LLM) – in this case OpenAI's GPT-4 accessed via an API. By testing their performance and then conducting manual (i.e., human) checks, we have found that LLMs such as GPT-4 can effectively place supermarket products into categories when given an appropriate prompt. Given a list of candidate categories and a product's name and original supermarket-defined category, the LLM can provide a suggested category designation.

38. This classification method is an iterative process. A stylised example is set out in Figure 4. To receive a response from the LLM, a prompt consisting of system instructions, a manual example, and a query are submitted. First, the system instructions define the task of categorisation, instructing the LLM to receive products and choose a new category from a defined list. Second, a manually defined example is provided, with both the query and assistant response written by a human. This is an example of prompt augmentation where already-answered prompts (queries) are included with the request to show how the LLM should respond (Liu et al., 2023).[4] Third is the query itself, with details of the products to be categorised structured identically to the provided example. In the example, just '6 Premium Seeded Batch Rolls' are categorised, but all supermarket products are processed in this way. Finally, in response to the prompt (containing the system instructions, the manual example, and the query), the LLM returns new categorisations. In this example, the rolls are placed into 'Bakery'.[5]

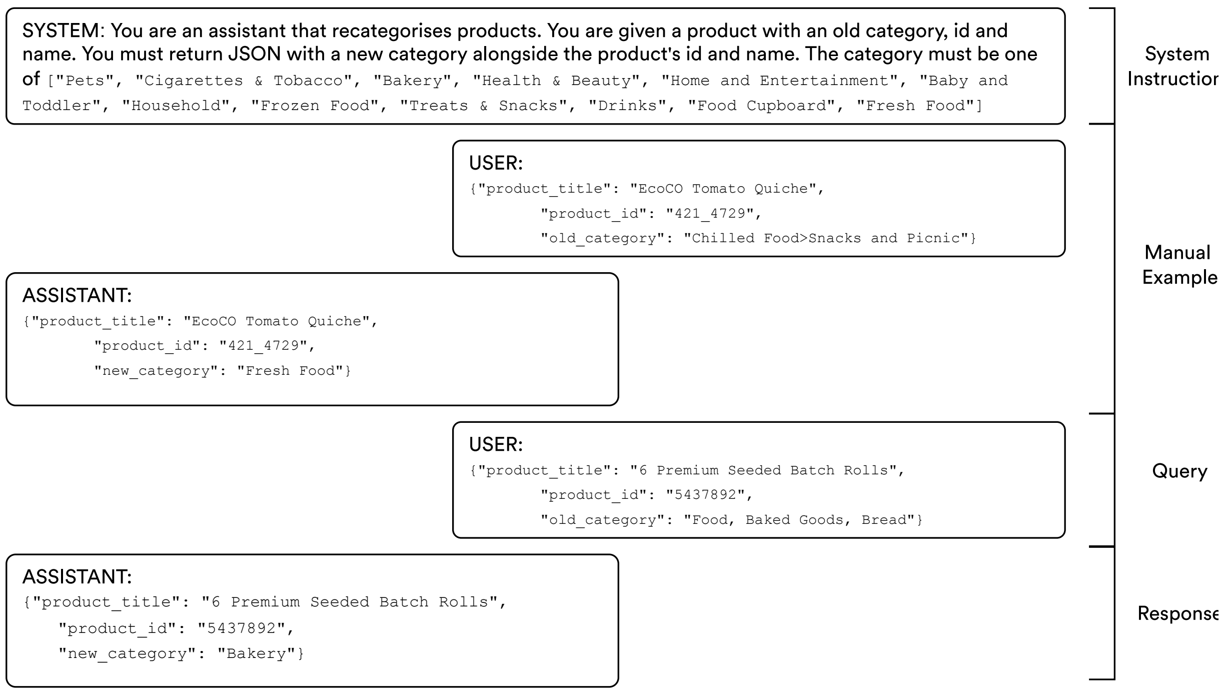
39. Following the final categorisation of each item, products given COICOP codes undergo a validation round. In batches of 20, the fit of supermarket products in their assigned categories are evaluated. We use specific prompts to test the LLM's assessments of whether products belong in their AI-chosen category. As well as a binary judgement, we also request a reason and confidence. In preliminary testing we observed that making the LLM first explain its reasoning led to better judgements, demonstrating the 'chain-of-reasoning' documented by Wei et al. (2023) and others. For example, the placement of 'Cashews and Raisins' into 'Fruit' in stage 1 returns a suggestion that this is a poor fit, along with the reason: 'Nuts are not fruit' (see the Annex for worked examples).

40. Our LLM process yields precise matches. Manual (human) evaluation of 400 products assigned COICOP codes (200 with the extra LLM validation stage, 200 without) shows that the categorisation algorithms perform strongly. For the products that were categorised using both the initial categorisation step and the LLM validation, 96% of assigned COICOP codes were correct. In contrast, for the products that only went through the first step of categorisation and did not receive LLM validation, the precision rate dropped to 76%. This significant decrease highlights the added value of the LLM validation in enhancing categorisation accuracy. These findings strongly indicate that the assignment of a COICOP code, especially when supplemented by LLM validation, is a reliable indicator of a product's true class.

[4] Augmenting in this way is particularly important for receiving properly formatted JSON responses as LLMs are adept at learning simple repetitive structures from prior examples.
[5] In further rounds, the rolls would be placed into subcategories, found within 'Bakery', until they have been assigned to a terminal node of our intermediate category system.

Figure 4: Example categorisation prompt and response

Categorisation of ‘6 Premium Seeded Batch Rolls’ into ‘Bakery’



Notes: Products are categorised multiple times, one layer of the intermediate system at a time - this shows the first layer. The prompt has been simplified; in deployment, extra instructions are given to ensure compliant formatting and products are categorised in batches of around 20. Source: Authors

41. While accurate, this process drops observations at each stage. The process begins with a set of around 140,000 products. Out of these, 128,000 are successfully categorised into a leaf node of our intermediate system, and 104,000 of them are assigned a COICOP code. The remaining products map to categories such as clothing and media. These remain in the raw dataset and may prove useful for future work. Since are focus here is food, they are dropped at this stage.

42. The final stage of the process discards a lot of data. Of the 104,000 products assigned a COICOP code, only 27,000 are confirmed to belong in their assigned category by the LLM validation step. This drastic reduction is not justified: human evaluation of the first stage of categorisation suggests that about 75% of the products are correctly categorised, but our LLM lets far fewer through. At this early stage in the project, our computerised research assistant is too zealous in its culling.

Stage 2 – Embeddings

43. The first stage of the matching process assigns supermarket products ('ASDA Basmati rice', for example) to relevant COICOP codes indicating a product's inclusion in a fairly broad class ('Bread', 'Poultry', 'Rice', etc.). The next stages involve going a layer deeper and matching products to the individual CPI items (e.g., 'SIX BREAD ROLLS-WHITE/BROWN', 'FRESH BONELESS CHICKEN BREAST' or 'BASMATI RICE 500G-1KG') that the ONS collects data on. To do so, we first generate a shortlist of possible candidates by matching CPI item names with similarly named supermarket products found in the same category. This is performed with the use of embeddings.

44. Embeddings are a way of representing high-dimensional data, such as text or images, in a lower dimensional space. They can reduce long strings of text to vectors (typically of 1000+ dimensions), while retaining much of the semantic meaning (OpenAI, 2023). Having a numerical representation of text then allow us to find the semantic similarity between any two strings. This ability to compute distances between any two strings (CPI items and supermarket products in our case) facilities the generation of our product shortlists: it is the supermarket products closest to each CPI item that form the given item's product shortlist. The combined use of embeddings and the COICOP categorisation stage enhances matching quality by eliminating poor matches as early as possible.

45. Figure 5 is a highly simplified representation of our process:

1. First, embeddings are found for the target CPI item ('Bread Rolls') and all other products with the same COICOP code (01.1.1.3 – 'Bread'). The figure shows a fictional two-dimension representation of these embeddings.
2. Second, the distances between every supermarket product ('EcoBread 600g Bloomer', ... , 'Cut-price White Baps') and the CPI item are calculated.
3. Finally, a candidate shortlist is formed by selecting the n (n=3 on the toy example) products that are closest to the bread rolls. In Figure 13, the white baps, sliced buns and grain rolls are closest to the 'Bread Roll' point, so these form the shortlist.

46. In reality, the process is more complicated. For example, 'text-embedding-ada-002' (the OpenAI model we use) does not have just two dimensions: it has 1,536. Despite this, the process follows the same steps:

1. Via the OpenAI 'ada-002' API, we fetch embeddings for every supermarket product and CPI item. We then normalise the embeddings to ensure we consider only distance, and not magnitude.
2. For each COICOP class, containing M CPI items and N supermarket products, we build a M×N distance matrix for each store, D, where $d_{m,n}$ contains the cosine similarity of CPI item m ($m \in \{1,2,...,M\}$) and supermarket product n ($n \in \{1,2,...,N\}$).

3. For each CPI item in the COICOP subclass, we take the 50 supermarket products with the highest cosine similarities[6] as our shortlist of candidates.

47. Incorporating embeddings into the matchings process offers several key advantages. It eliminates the need for manual definition of substrings to match by (Breton et al., 2015), simplifying the process significantly. Additionally, it excels in understanding semantic meanings, not just superficial term similarities. Altogether, it acts as an efficient and low-cost initial stage in our classification process.

48. As a lower-complexity filtering stage, embeddings-driven classification is not without limitations. In some cases, embeddings can yield misleading similarities. For example, considering just cosine similarity, 'Butternut Squash' is more closely related to 'Fruit Squash' than 'Pumpkin'[7]. Additionally, the matchings can struggle to accurately represent consumer spending; There exists a danger of matching to products that are semantically the closest to a matched term but not as commonly purchased. For example, own-brand chocolate biscuits are closer matches to 'chocolate biscuits' than branded items (e.g. 'Chocolate Digestives') which are more representative of consumer preferences.[8] To address these issues, we have begun defining and inserting common brand names into our terms for embedding. Furthermore, we cast our net far during the initial embeddings stage, yielding around 350 potential matches for each cpi-item, ready to be refined and culled through our final stage: LLM driven assessment.

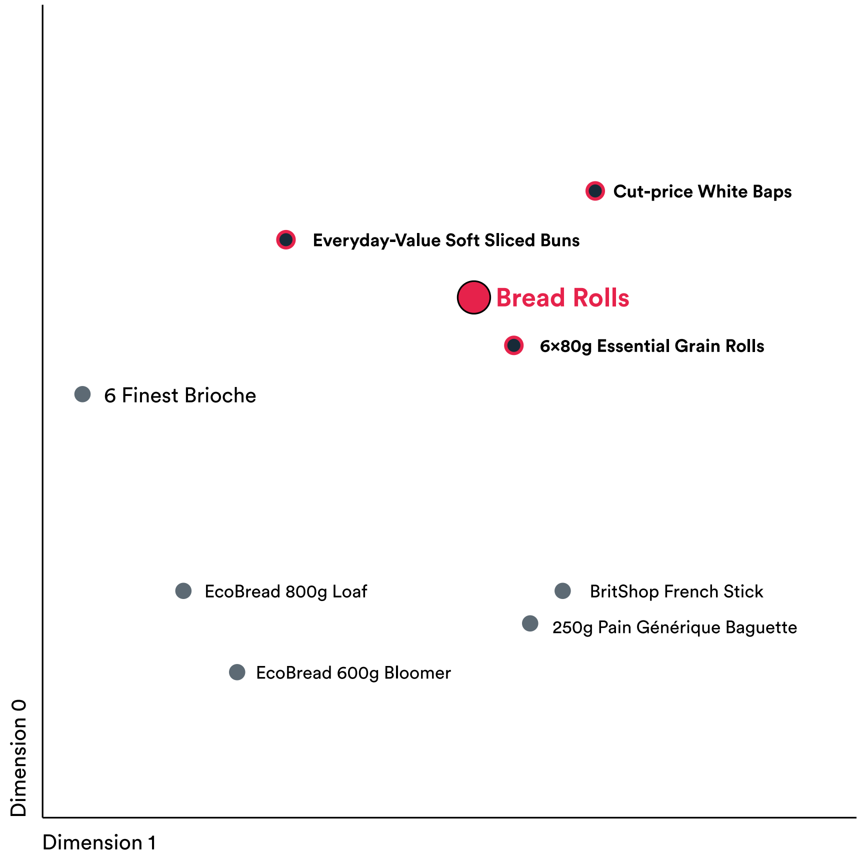
[6] Cosine similarity being positively related to the similarity of two vectors.

[7] Barely; Cosine similarities of 0.92 and 0.90 respectively.

[8] We are grateful to Liam Greenhough for his helpful guidance around this issue.

Figure 5: Forming a candidate shortlist with embeddings

A toy example



Notes: This shows a highly simplified representation of our use of embeddings to generate a product shortlist for a CPI item (Bread Rolls). Source: Authors

Stage 3 – Final classification

49. The third stage of our matching process involves selecting just the best products for each CPI item from the candidates found using embeddings in the previous section. Again, using the GPT-4 API, we present a CPI item to test for and sets of 20 candidate supermarket products. In response to a carefully constructed prompt, the LLM then returns assessments, with reasoning and confidence scores, of whether each supermarket product is a match for the CPI item.

50. As with the COICOP validation, a system prompt defines the task to be performed and manually completed examples (omitted for brevity) show a fully worked example to reinforce formatting expectations. In the query, an item to test for and supermarket products from the embeddings-generated shortlist are given in a standard JSON structure. In

response, the LLM returns JSON indicating whether each product is a good match for the candidate item.

IV Results

51. The dataset has several uses. For example, it can be further linked with trade data in order to establish the geographical source of inflation (Davies and Hellings, forthcoming). Here we focus on two literatures: first, the measurement of nominal rigidities, an important input into macroeconomic models; and second, the development of automated consumer price indices (AutoCPIs).

Nominal rigidities

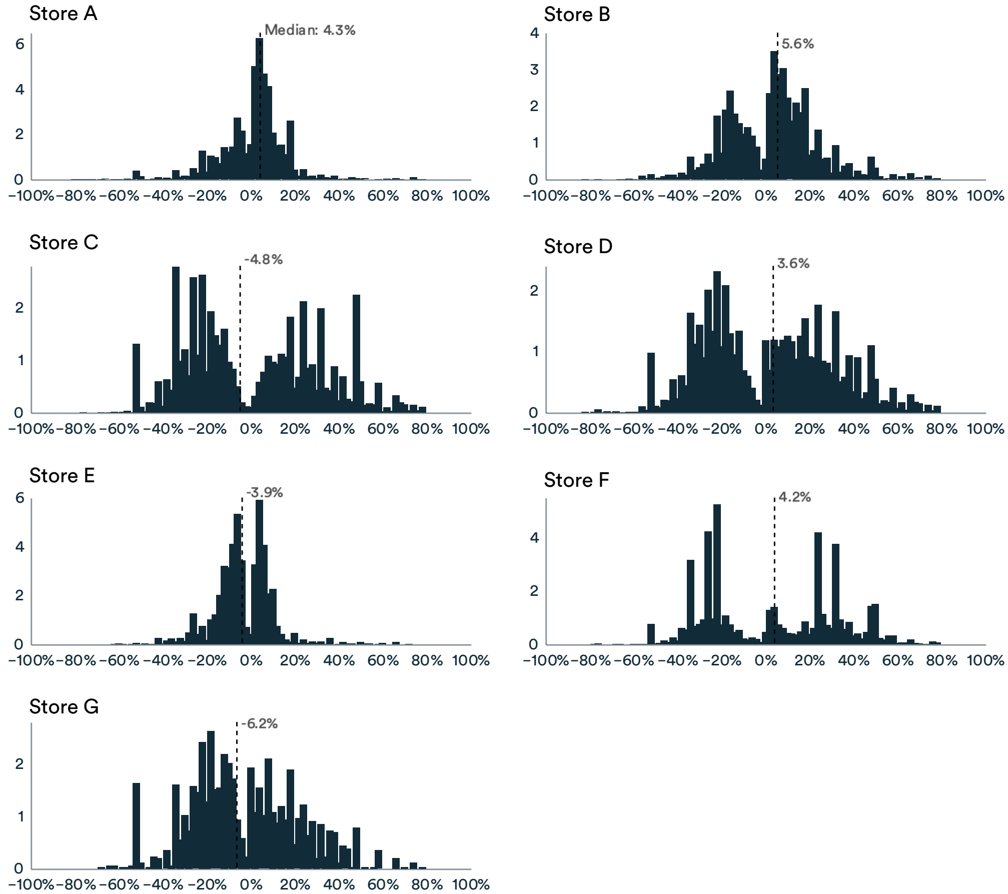
52. This section presents evidence on the frequency, size, and timing of price changes. These observations are helpful as they are indirect evidence of the underlying rigidities that affect firms' price-changing behaviour and, at an aggregate level, govern inflation. Intuitively, in a frictionless competitive world where price changes are costless, the distribution of prices would be smooth and price changes frequent. Examining the distribution of prices, as well as the schedule and size of price changes, can help us determine the rigidities and models most suitable for the UK grocery market.

53. If changing prices is costly then small tweaks to prices should be rare. In traditional menu costs models, firms avoid making small adjustments, making large price changes when the optimal prices diverge significantly from current prices (Sheshinski and Weiss, 1977). The UK data point to firm-level heterogeneity here (Figure 6). For example, store E makes lots of small changes of size ±4%; store F makes few small price changes, with most being around ±20%.

54. The dearth of small price changes is an indication of 'pent up' desire to change prices. Adjustments are not made until prices have drifted far enough away from their optimal level for it to be worth bearing the cost of changing them (ibid.). Most firms avoid the smallest price changes, leaving a depression in most distributions around 0%. However, this is not found for every store in our dataset. Store A is atypical, with a uni-modal distribution centred around 4% and no pronounced lack of small price changes.

Figure 6: Store Price Change Distributions

July 2023 – January 2024



Notes: Includes all price changes, except those identified as anomalous, of all items, not just CPI item equivalents.
Source: Authors' calculations. Anonymised store names are non-constant across figures.

Frequency and timing

55. In the daily data, most prices do not change. After seven months of data collection, we have 19.1 million observations where there is a comparator on the previous day. Of these, 1% represent price changes. Using a week as our window, 5% of prices change. Over a month our measure is 16% which is reassuringly close to that seen in official ONS microdata (Davies, 2021). On the longest run (six months) of 142,000 shop-item pairs, close to 44% experience at least one price change, with 34% ending the period on a different price to their initial price in the dataset. Table 4 below sets out price changes at various different time horizons.

Table 4: Price change frequency, various windows		
Window	Comparable prices	Price changes
Daily	19.1m	1%
Weekly	18.3m	5%
Monthly	16.2m	17%
3 months	11.0m	24%
Long run (6m)	143k	34%

Notes: The windows used in the first four rows are 1, 7, 28 and 84 days respectively.

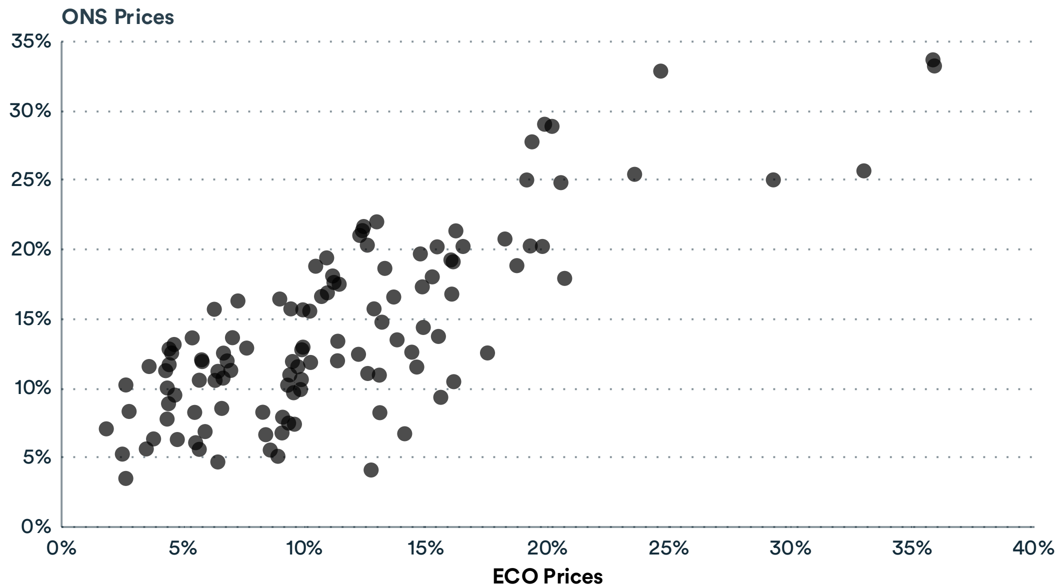
Source: Authors' calculations.

56. Prices change every day at our retailers. That said, there is a distinct weekly pattern, with prices collected on a Monday representing 46% of all price changes, suggesting Sunday is the main price change day. Around 20% of prices have changed by Tuesday. Few new prices are picked up on Wednesdays, Thursdays, or Fridays. This pattern likely reflects the main price changes undertaken physically in-store. Early store closure, in accordance with Sunday Trading Law, provides an opportunity for staff to manually change prices. This suggests in-store prices represent a constraint on whether on-line prices can change.

57. We also examine the frequency of price changes at the item level (Figure 7). The most flexible prices over the six months we investigate are oils and drinks. The data we match to the item 'OLIVE OIL' change prices 36% of the time in our data over a monthly window. This compares to 33% in the official ONS microdata. As a sense check on our prices, we compare item-level flexibility with that seen in the ONS data. The correlation between our data—still at an experimental stage—is 0.8.

Figure 7: Price flexibility: official data and our data

Frequency of price changes, by CPI item, monthly, past six months



Notes: 0.8 correlation coefficient. Shows the probability of a product changing price each month. Includes all products classified as a CPI-item equivalent. Source: Authors' calculations; ONS

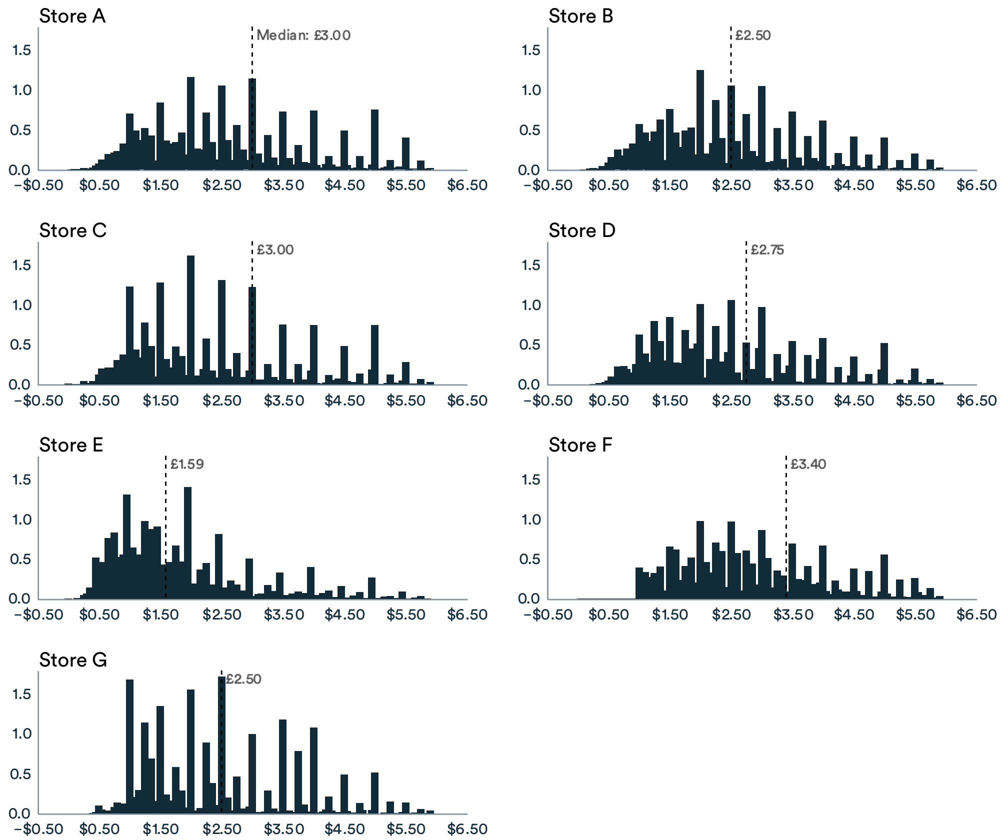
Anchor prices

58. Retailers usually set prices at common anchor points. Recent micro-data research (Davies, 2021; Knotek, 2010) has found that most prices end in just a few subdivisions (.99, .50 and whole prices). This behaviour can be explained with the concept of charm pricing and left-digit bias (Thomas and Morwitz, 2005), where firms strategically set prices to make them seem lower to consumers. Our dataset allows us to further study these well-established pricing practices.

59. A majority the prices in our dataset end in .00 (30%), .50 (16%) or .25 (6%). Figure 8 shows the price distributions of each firm in our dataset. We observe clear peaks in the distributions around anchor points, with few prices set while irregular prices are avoided. For example, just 4% of prices end in the digit 4. Though our data shows abundant use of other anchor prices, we find just 9% of prices end in the digit 9 – far fewer than Knotek (2010), who observes this for 62% of prices. Our finding is consistent with Davies (2021), which describes a decline in the use of .99 prices which can possibly be explained by the shift towards online shopping.

Figure 8: Price distributions

Histograms with 5p bins



Notes: Excluding the top 10% of prices (approximately >£10.50).Source: Authors' calculations.

Synchronisation

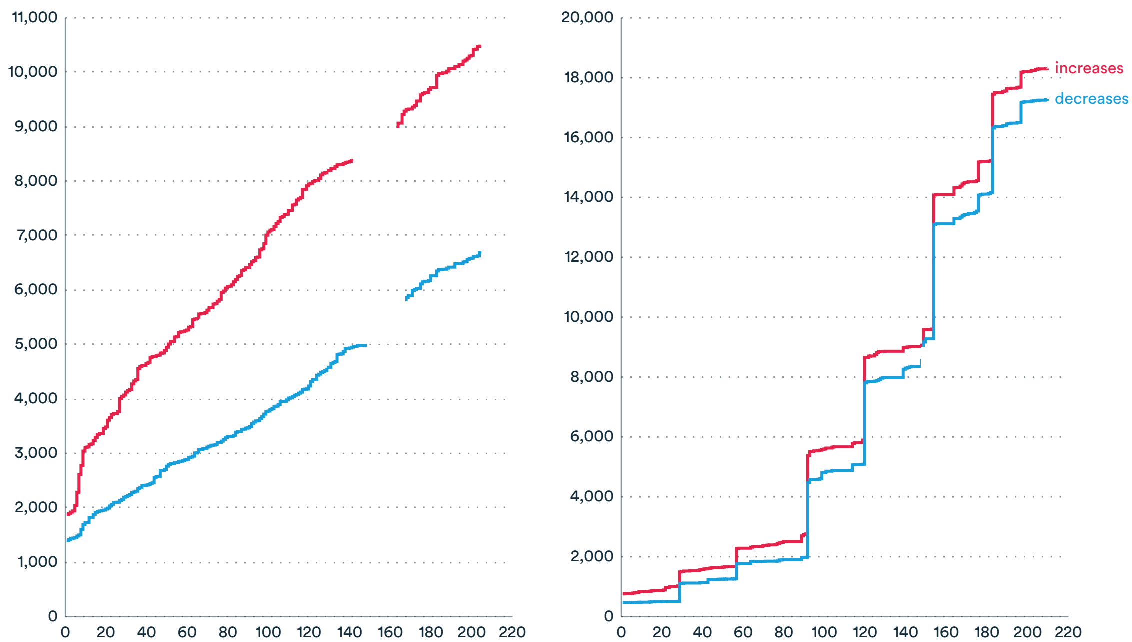
60. Another feature we observe—synchronisation—points towards a menu cost as being the ultimate source of rigidity. Rather than changing prices continuously, firms tend to adjust many prices at once, often at regular intervals, in response to changing conditions (in time dependent models such as Calvo, 1979) or the costs of repricing (in modern menu costs models, such as Kehoe and Midrigan, 2015). In recent years, within-firm synchronization has been observed in Denmark (Dedola et al., 2019), Israel (Bonomo et al., 2022), and Norway (Nilsen et al., 2021).

61. We observe differing degrees of synchronisation across firms. Figure 9 shows price changes in one indicative month. Some retailers 'synchronise' their prices, with a sudden rush of re-pricing activity on a given day. Others seem to have menus that are in a constant flow, with no temporal alignment of price activity. In Figure 9 we can see the two extremes; one retailer (store B) has made most changes on just a few days, with some days seeing a

full 10% of prices changed. The other retailer (store A) has instead made many small adjustments.

Figure 9: Price synchronisation

Cumulative price changes since august 2023, by direction. Store A (LHS) and B (RHS)



Notes: Excluding price changes identified as anomalous. Anonymised store names are non-constant across figures. Source: Authors' calculations.

62. The starkly different behaviour of these firms seems to suggest that there is heterogeneity in the menu costs they face. This step changes of B are indicative of a 'economy of scope' in price changing: once a firm has paid some fixed cost to engage in price changing, it is then optimal to change lots of prices on that day (Kehoe and Midrigan, 2015). Changing prices in brick-and-mortar stores fits this model intuitively. Changing a single price is time-consuming due to the fixed costs in receiving updates and preparing price 'tickets', but it becomes more efficient when updating multiple prices at once, as most time is spent navigating the store rather than on the individual changes.

AutoCPIs

63. Using our dataset, we can produce automated estimates of grocery price inflation. Prices, categorised according to the LLM-driven methodology presented in this paper, are

used to form baskets of supermarket goods. These are weighted according to ONS COICOP subclass and item weights to produce two AutoCPI estimates of grocery price inflation. These balance breadth with congruency with official figures and are defined as follows:

- **CPI matching measure (narrow)**. Our primary estimate, here we match and weight at the CPI item level, incorporating around 150,000 prices per month (5,000 per day) for 178 items in the CPI basket. This compares with the 47,000 prices the ONS collects each month for the same items.

- **Expenditure matching (broad).** Here we are matching and weighting at the COICOP sub-class level ('Bread', 'Poultry', 'Rice', etc.), and are matching around 400,000 prices per month.

64. Our AutoCPI incorporates products from the COICOP division, 'Food and Non-Alcoholic Beverages'. Since mid-2021, this division has seen significant price growth, with year-on-year inflation peaking at 19.1% in March 2023. However, within this period, monthly inflation has been highly volatile, with month-on-month inflation falling from highs of 2.2% in July 2022 and 2.1% in February 2023, to a low of -0.4% in January 2024.

65. Consistent with official index calculation (ONS, 2019), we employ two primary index methods, Jevons and Laspeyres. At our lowest strata, we calculate Jevons indices. We compare the geometric means of prices ($\mathrm{p_i}$) for identical sets of products ($\mathrm{i} = [0, \dots, \mathrm{n}]$) across stores collected in one period (t[9]) to another (0):

$$\mathrm{I}^{0,\mathrm{t}} = \frac{\sqrt[\mathrm{n}]{\prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p}_\mathrm{i}^\mathrm{t}}}{\sqrt[\mathrm{n}]{\prod_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p}_\mathrm{i}^0}}$$

66. In contrast to official CPI calculation, we can uniquely identify products at the SKU level[10] and possess a weight for every store. This would predispose our elementary index calculation towards a Laspeyres index with arithmetic means of weighted observations. However, the Jevons indices are instead calculated for congruency with official calculation and because of a closer fit in testing. To weight our observations, we employ replication factors similar to those computed for centrally collected ONS prices, where observations are repeated proportionally to their store's market share.

[9] The index day, on which the ONS collects prices, lies around the 9-18th of the month and is announced only after price collection. We assume an index day of the 14th each month.
[10] We treat all SKUs as distinct and currently do not capture product relaunches, for example changes in quality or size. We are exploring methods, engaging existing literature (Sands, 2021), to associate and quality-adjust relaunches with the products they replace.

67. To aggregate our item indices, we use standard Laspeyres indices. Baskets of 178 CPI items (narrow AutoCPI) and 54 COICOP sub-classes (broad AutoCPI) (j) are brought together with standard, ONS provided, CPI weights (w_j):

$$AutoCPI_t = \sum_{j=1}^{J} w_j I^{0,t}$$

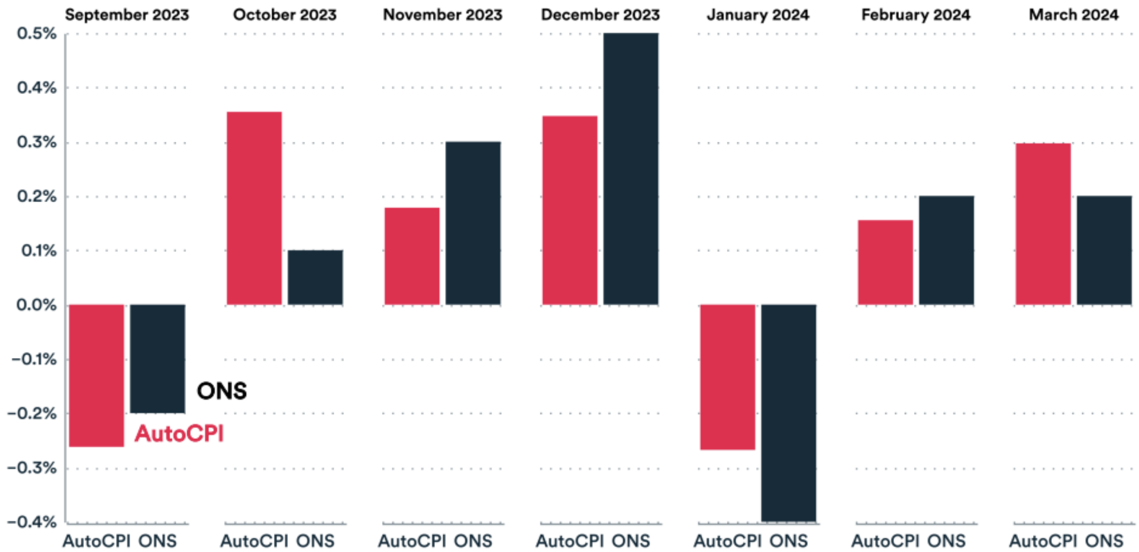
We primarily calculate indices comparing one month to the one before. If chained together, these estimates can suffer from chain drift over time, where changes in the makeup of the basset as products enter and dropout of the market lead to divergence from standard fixed basket indices (ILO, 2004, 1.46-1.51). As our aim is to provide early, and current, estimates of inflation since the last official release rather than replace traditional collection and calculation, we partially avoid these issues. However, we are keen to strengthen our index calculation and are exploring alternative index methods.

The item-matching AutoCPI

68. Our primary measure is the CPI-matching narrow AutoCPI. Price observations identified among 187 specific CPI items are first aggregated using Jevons indices, and then a single index is constructed with the Laspeyres formula, as described above. This selection encompasses the majority of the 210 CPI items found in the 'Food and Non-Alcoholic Beverages' ONS basket, accounting for 77% of its total weight. A total of 23 items not typically sold in supermarkets (primarily restaurant food and takeaways) are excluded.

69. To aid representativeness, supermarket price observations are filtered with comparisons to ONS microdata. During the classification process, the size of products, such as the number of teabags per box, is initially disregarded which can lead to subtle mismatches. For example, a price for 80 tea bags could be erroneously included in the 'TEA BAGS PACKET OF 210-240' item. To address this, we compare price observations to the distribution of ONS price microdata and eliminate prices that fall outside the 95 central percentile of ONS prices. The filtering criteria vary. For some items, such as 'POTATOES-NEW-PER KG', unit prices are used, while for others, like 'FRESH VEG-CAULIFLOWER-EACH', the sticker price is considered. This filtering serves a dual purpose: it helps eliminate mismatches due to incorrect sizes and provides an additional layer of protection against misclassification of products.

Figure 10: CPI-matching AutoCPI estimates
Mid-month to mid-month changes, AutoCPI vs ONS



Notes: This figure compares estimated AutoCPI indices to real ONS item indices since September 2023.
Sources: Authors' calculations, ONS.

70. By refining our dataset through price filtering, we establish a stronger correlation between the mean prices of our web-scraped data and the ONS microdata Initially, mismatches such as the inclusion of single item products in items meant for multipacks (such as 'COLA/FIZZY DRINK 330ML PK 4-8') distort the accuracy of our findings. However, the application of price filtering effectively eliminates these major inaccuracies. Despite these improvements, some discrepancies persist, particularly in ONS categories with broad specified ranges, such as 'JOINT OV/READ GAM/POR 450-900G'. This suggests a potential bias in our data collection towards items at one end of the specified range.

71. The initial six-month results from our narrow AutoCPI approach are promising, despite the limitations of a short back series. Our estimates have been proven directionally accurate each month. In the best performing months, September 2023 and January 2024, there are just 0.02pp and 0.08pp spreads, respectively.

The broad AutoCPI

72. The affordability of online data collection facilitates experimentation with the composition of our basket. Traditional price collection from brick-and-mortar stores is expensive and labour-intensive, whereas our marginal cost of collection is close to zero.

Consequently, we are able to compile a broader, alternative AutoCPI that encompasses all supermarket products categorised under a COICOP subclass, not just those directly matching CPI item criteria. This means we include products with a cumulative 400,000 prices per month, allowing for a comprehensive investigation into how the scope of product inclusion affects inflation statistics.

73. Broader CPI measures have the potential to capture price movements overlooked by official baskets. For instance, our Broad AutoCPI for September showed a price increase, contrary to the price decrease suggested by both official statistics and our narrow AutoCPI. Investigating the items included in the 'broad' but excluded from the 'narrow' AutoCPI sheds light on this discrepancy. Notable examples include substantial price hikes in premium grain snacks, such as Ryvita and snacking poppadoms, which do not fit into the 'CREAM CRACKERS PACK 200G-300G' CPI item category. However, caution is warranted when interpreting these results, as the broad AutoCPI's lack of item-level weighting introduces greater volatility to the data.

74. AutoCPI estimates are high resolution, allowing us to observe price changes at daily, not monthly, intervals. For example, official figures and AutoCPI estimates for January 2024 both indicate a similar overall price change (-0.037% and -0.039%, respectively), but using traditional data, we have no idea of the path prices took in the interim. Our data reveals that for January 2024, the majority of price adjustments occurred shortly after Christmas, with prices nearly flat until then. This detailed observation capability provides valuable insights for other areas of research, such as the forthcoming research on responses to non-tariff barriers by Davies and Hellings (forthcoming). Additionally, while food inflation is relatively low at present, in times of crisis daily observations can reveal price shocks long before they are incorporated into official statistics.

V Conclusion

75. This paper contributes a new dataset of British grocery prices, introduces a new methodology for efficient and accurate LLM-driven product classification, examines pricing behaviour in the UK grocery market and provides early estimates for intra-month grocery price inflation.

76. The dataset provides insights into prices at an unprecedented frequency and granularity. By collecting prices daily, we are able to examine price movements with more

resolution than the current standard, investigating the timing of changes and facilitating daily CPI estimation.

77. Our classification system uses recent advancements in Large Language Models to make accurate matches while requiring little labour. Combining multiple layers of LLM assessment and re-assessment, we correct for errors and achieve high precision.

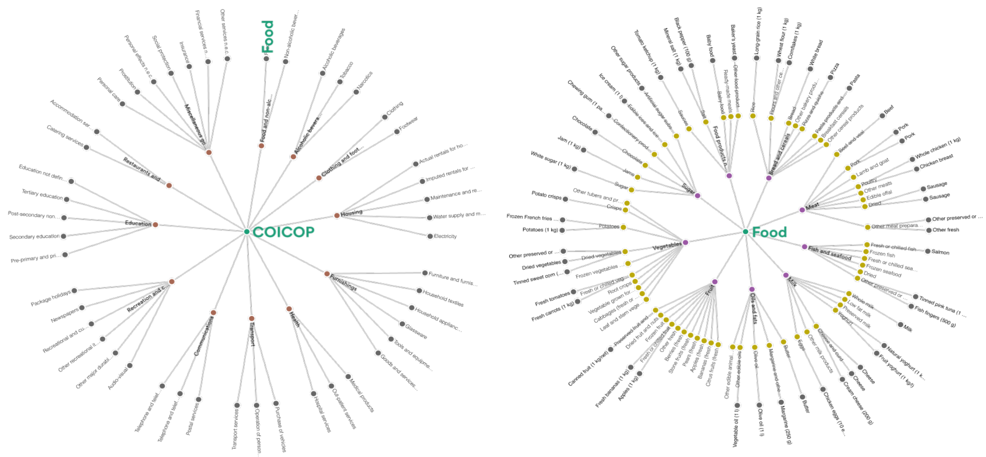
78. With data sourced directly from supermarket websites, we can examine pricing at the store level, not just at the item or industry level. This approach reveals heterogeneity in how and when firms change prices, offering a more nuanced understanding of inflation and its underlying causes.

79. In addition to our research contributions this paper has significant policy relevance. We have shared our initial findings and methodology with colleagues at the Bank of England and ONS, and once our results have been peer reviewed and code verified, we intend to make both our code and our daily data-set available to policymakers.

Annex: Additional Figures

Figure 11: The COICOP System

Tree diagrams showing the first 2 divisions of the COICOP system (left) and all subdivisions of ’Food’



Notes: Most other branches on the left diagram have subdivisions of similar depth and breadth to ’Food’ but are omitted for clarity. Sources: Authors’ illustration, ONS.

Figure 12: Category Validation

Assessing assignment with an LLM

System Instructions

SYSTEM: You are an assistant that checks whether products belong in a category. You receive a category name and a list of products. You need to decide whether each product belongs in the category with 20 words of reasoning and give a confidence from 0 to 1. It is important you only return JSON in the form of [...].

Manual Example

ASSISTANT:
```
{"product_title": "EcoCO Tomato Quiche",
```
Truncated for Brevity

Query

USER:
```
{"candidate_category": "Fruit",
    "products": [{"id": "646762804244804245",
         "product_title" : "Tesco Cashew Nuts & Raisins",
         "product_breadcrumb": "Cupboard', 'Dry Fruit &
Nuts"}]}
```

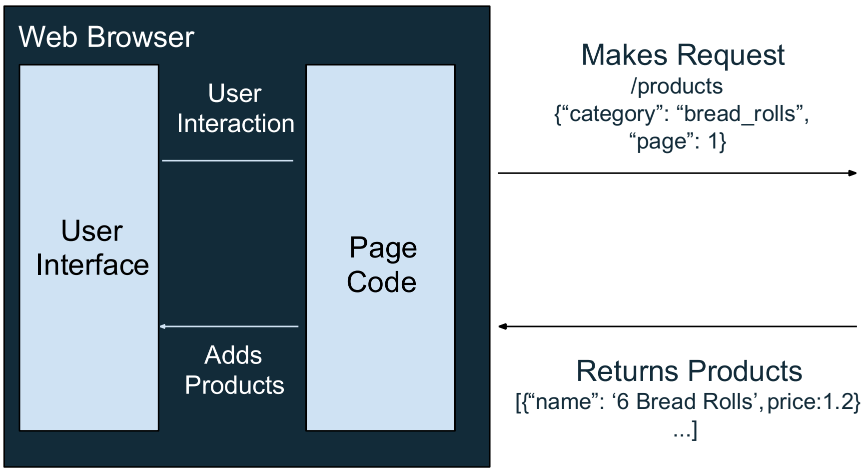
Response

ASSISTANT:
```
{"category_name": "Fruit",
    "products": [{
        "id" : "646762804244804245",
        "product_title" : "Tesco Cashew Nuts & Raisins 200G",
        "candidate_category": "Fruit",
        "reason": "Nuts are not fruit",
        "score": 0.4,
        "belongs" : false}]}
```

Notes: In production much larger batches are run. Source: Authors’ illustration

Figure 13: AJAX Requests

A supermarket requesting new products



Notes: The request and response structure are highly simplified. Typically, the request body will include additional metadata and cookies.

Annex: Collecting Data Responsibly

1. The Economics Observatory data team are working on several projects that aim to improve UK economic statistics. In doing this we follow university research ethics guidance, and adopt best practices proposed by the ONS and ECB. A full paper on this issue is available as a separate document (here).

2. In summary:

- **Minimising Burden:** Our collection makes up a very small proportion of the total visits retailers receive as prominent national companies. We further minimise the burden by scraping outside of peak times (starting at 6AM) and by requesting just the resources we need. Where possible, we request only the product information directly, reducing the quantity of resources downloaded by up to 80%.
- **Robots Exclusion Protocol:** Each site's robots.txt sets out webpages that should not be accessed by automated systems. We abide by these, avoiding pages listed. The robots.txt do not usually list non-webpage resources like the product information we load through XHR requests (our second approach), but we still follow the robots.txt in spirit by avoiding resources requested by pages listed in the robots.txt.
- **Privacy:** We do not, and can not, collect user identifying information such as personal details that would fall under GDPR. All the data we collect are publicly available product information, accessible without logging in or non-standard behaviour.

3. We are also careful to follow best practises in data transformation and dissemination:

- **Anonymisation**: When sharing research findings, we will not identify supermarkets directly when discussing price dynamics. Supermarket identities will only be discussed in the context of standard summary statistics. To protect firms' commercial interests, only aggregate price movements will be shared.
- **Security:** All our underlying data is stored securely, through standard encrypted cloud infrastructure providers. We are not sharing non-aggregated data or findings outside of the LSE and have no commitments to do so. We would like to share the data with three parties: HM Treasury, The Bank of England, and the ONS, where we think there is a clear policy interest.

4. Finally, we note that there is a clear public interest motivating our data collection. There exists an important debate about the quality and future of traditional data sources. See e.g. Bernanke (2024) and Devine (2023).

References

Bank Underground. (2024, January 17). “Beyond the average: Patterns in UK price data at the micro level by Lennart Brandt”, Natalie Burr, and Krisztian Gado. Inflation, Microdata, Monetary Policy. Available at https://bankunderground.co.uk/2024/01/17/beyond-the-average-patterns-in-uk-price-data-at-the-micro-level

Beer, C. Ferstl, R. Graf, B., & Rumler, F. (2023). “Price setting of groceries in times of high inflation: What to learn from webscraped data?”. Oesterreichische Nationalbank.

Bernanke, B. (2024, April). “Forecasting for monetary policy making and communication at the Bank of England: a review”. Bank of England Independent Evaluation Office.

Bonomo, M. Carvalho, C., Kryvtsov, O., Ribon, S.; & Rigato, R.D. (2022). “Multiproduct Pricing: Theory and Evidence From Large Retailers”.

Breton, R., Clews, G., Metcalfe, L. Milliken, N., Payne, C., Winton, J. & Woods, A. (2015, September). “Research Indices Using Web Scraped Data”. Office for National Statistics. Accessed: 2024-01-09. Available at: http://www.ons.gov.uk/ons/rel/cpi/consumer-price-indices/research-indices-using-web-scraped-price-data/index.html.

Cavallo, A. (2013). “Online and official price indexes: Measuring Argentina’s inflation”. Journal of Monetary Economics, 60(2), 152–165. doi:10.1016/j.jmoneco.2012.10.002

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

Competition and Markets Authority. (2014). “Review of the Pricing Practices Guide: Response to the Call for Evidence by the Trading Standards Institute”. Available at: https://assets.publishing.service.gov.uk/media/5a7db5b2e5274a5eb14e6b72/CMA_response_to_TSI_Call_for_Evidence_on_PPG.pdf [Accessed 20 Mar. 2024]

Davies, R. (2021, February). “Prices and inflation in the UK - A new dataset”. CEP Occasional Papers 55. Centre for Economic Performance, LSE.

Dedola, L., Kristoffersen, M. S., & Züllig, G. (2019, September 19). “Price synchronization and cost pass-through in multiproduct firms: Evidence from Danish producer prices”. European Central Bank (ECB).

Devine, B, “Has labour market data become less reliable*?”* House of Commons Library.

Eugster, P. & Uhl, M. W. (2024). "Forecasting inflation using sentiment". Economic Letters.

European Parliament and Council of the European Union, 2019. Directive (EU) 2019/790. Available at: https://www.legislation.gov.uk/eudr/2019/790/contents

Eurostat. (2016). "European Classification of Individual Consumption According to Purpose (ECOICOP)". Accessed: 2 January 2024. Available at: https://showvoc.op.europa.eu/#/datasets/ESTAT_European_Classification_of_Individual_Consumption_according_to_Purpose_%28ECOICOP%29/metadata.

Horrigan, M. W. (2013). "Big Data: A Perspective from the BLS". Amstat News, January, pp. 25-27.

International Labour Office. (2004), Available at: https://webapps.ilo.org/CPI/doc/cpi_manual_en.pdf

Kantar. (2024). "Grocery Market Share Great Britain". Available at: https://www.kantarworldpanel.com/grocery-market-share/great-britain/snapshot. Accessed: 31 January 2024.

Kehoe, P. & Midrigan, V. (2015). "Prices are sticky after all". Journal of Monetary Economics, 75, pp. 35-53

Knotek, Edward S. (2010). "The roles of price points and menu costs in price rigidity". Research Working Paper RWP 10-18. Federal Reserve Bank of Kansas City.

Liu, P. Yuan, Weizhe. Fu, Jinlan; Jiang, Zhengbao; Hayashi, Hiroaki; Neubig, Graham. (2023). "Pre-Train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing". ACM Comput. Surv., 55(9). DOI: 10.1145/3560815.

LSE (2024, April). “Ethics review considerations: A quick guide for researchers (students or staff)”. Available at https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethicsReviewConsiderations-QuickGuide.pdf

LSE (2024, February). “Online ethics review form – questions only”. Available at https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-review-form-questions-only.pdf

Macias, P., Stelmasiak, D., & Szafranek, K. (2023). “Nowcasting food inflation with a massive amount of online prices”. In: International Journal of Forecasting, 39(2), pp. 809–826.

Metcalfe, E., Flower, T., Lewis, T., Mayhew, M., & Rowland, E. (2016, November). “Research Indices Using Web Scraped Price Data: Clustering Large Datasets into Price Indices (CLIP)”. In: Office for National Statistics (ONS).

Monroe, J., (2022, January), “We’re pricing the poor out of food in the UK – that’s why I’m launching my own price index”, The Guardian

Nakamura, E., & Steinsson, J. (2008, November). “Five Facts about Prices: A Reevaluation of Menu Cost Models”. The Quarterly Journal of Economics, 123(4), pp. 1415–1464.

Nakajima, J., Yamagata, H., Okuda, T., Katsuki, S., & Shinohara, T. (2021, October). “Extracting Firms' Short-Term Inflation Expectations from the Economy Watchers Survey Using Text Analysis”. Bank of Japan Working Paper.

Naylor, J., et al. (2014). “The ONS Big Data Project”. Office for National Statistics.

Nilsen, Ø. A., Skuterud, H., & Munthe-Kaas Webster, I. (2021). Price Change Synchronization within and between Firms. NHH Dept. of Economics Discussion Paper No. 15/2021. Norwegian School of Economics (NHH) - Department of Economics; IZA Institute of Labor Economics

Office for National Statistics (ONS). (2020, September). “New index number methods in consumer price statistics”. In: Office for National Statistics.

Office for National Statistics (ONS). (2020b). “Web scraping policy”. Available at https://www.ons.gov.uk/aboutus/transparencyandgovernance/datastrategy/datapolicies/webscrapingpolicy

Office for National Statistics (ONS). (2023). “Research and developments in the transformation of UK consumer price statistics”: December 2023.

Sands, H. (2021). ‘Research into the use of scanner data for constructing UK consumer price statistics’. Office for National Statistics

Stelmasiak, D., Szafranek, K., Macias, P., & Błażejowska, A. “Online food prices and shocks to product availability since Covid-19”. Narodowy Bank Polski.

Sheshinski, E., & Weiss, Y. (1977). Inflation and Costs of Price Adjustment. The Review of Economic Studies, 44(2), 287–303.

Thomas, M., & Morwitz, V. (2005). “Penny Wise and Pound Foolish: The Left-Digit Effect in Price Cognition”. Journal of Consumer Research, 32(1), 54-64.

United Nations Statistics Division. (2018, December). “Classification of Individual Consumption According to Purpose (COICOP) 2018.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2023). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv: 2201.11903

What are embeddings? (2023). OpenAI. Available at: https://platform.openai.com/docs/guides/embeddings

World Wide Web Consortium. (2018, June). WebDriver. W3C Recommendation. Available at: https://www.w3.org/TR/webdriver1/.

Economics Observatory

Economic and Social Research Council

www.economicsobservatory.com