Immeasurability of shrinkflation in the CPI? Automatic downsizing detection using scanner data

Jacek Białek, Anna Bobel, Dagmara Oprych-Franków*
Department of Trade and Services, Statistics Poland

18^{th} Meeting of the Ottawa Group, Ottawa, 13 - 15 May, 2024

April 24, 2024

Abstract

Measuring price changes is increasingly associated with "hidden" inflation, when product prices remain seemingly unchanged, but due to the reduction in the content and size of packaging, consumers pay more per unit of product. Reducing packaging, not only in the group of goods and food products, is a process of reducing the quantity (weight) of products and sometimes even changing their composition or lowering quality while maintaining the prices of these products at the current level. It sometimes happens that reducing the quantity of a product in a package is accompanied by an increase in price, even though, as research shows, consumers are more discouraged by price increases than by reductions in package size. The practice of shrinkflation negatively affects consumers' ability to make informed purchasing choices and also affects the credibility and perception of CPI.

A big challenge for public statistics is the detection of this type of practices, and therefore reliable verification of a sample of products and automation of the selection of products generating shrinkinflation. This work presents the results of the authors' research on the phenomenon of shrinkinflation withe the use of scanner data. The paper proposes an automatic procedure for selecting products affected by downsizing and presents the results of the analysis of this phenomenon for selected food and non-food groups of scanned products.

JEL codes: C43, C63, C81, D12
Keywords: inflation measurement, scanner data, shrinkflation, downsizing, price indices

*This paper represents the subjective opinion of authors and does not represent position of Statistics Poland on the content presented.

1 Introduction

In economics, *shrinkflation*, also known as *package downsizing* or *weight-out*, is the process of items shrinking in size or quantity while the prices remain the same. The first use of the term *shrinkflation* with its current meaning has been attributed to the economist Pippa Malmgren. Brian Domitrovic (2009) used the term *shrinkflation* to contrast with *stagflation* (which is persistent inflation combined with stagnant consumer demand and relatively high unemployment).

Skimpflation involves a reformulation or other reduction in quality. The term *shitflation*, which refers to the case of maintaining the price of a product while reducing quality, can also be found in the literature (Giner 2022). This term has already been used by online communities such as *Reddit* or *Twitter* since 2020. A more precise economic definition can be presented as follows: *shrinkflation* is a rise in the general price level of goods per unit of weight or volume, brought about by a reduction in the weight or size of the item sold.

A change in the size of a product in the absence of a change in price can be difficult to perceive for the consumer, who sometimes may not even notice that it is a *de facto* change to his or her disadvantage. The natural question that arises is how widespread this phenomenon is and what is its impact on inflation measurement. For instance, the *UK Office for National Statistics* wrote in 2019: [1] "We identified 206 products that shrank in size and 79 that increased in size between September 2015 and June 2017. (...) The majority of products experiencing size changes were food products and in 2016, we estimated that between 1% and 2.1% of food products in our sample shrank in size, while between 0.3% and 0.7% got bigger.". Examples of downsizing, which involves products from well-known brands, are presented in Table 1.

Table 1: Examples of downsized products

Company	Product	Date of size reduction	Size reduction	Size reduction [%]
Nestlé	After Eight	2010	From 200 g to 170 g	15
Burton' Biscuit	Cadbury Fingers	2015	From 125 g to 111 g	11.2
Milka	Milka Nuts & Raisins	2017	From 300 g to 270 g	10
Unilever	Ben & Jerry ice-cream	2020	From 500 ml to 465 ml	7
General Mills	Boxes of cereal	2021	From 19.3 oz to 18.1 oz	6.2
Procter & Gamble	Roll of toilet paper	2022	From 264 to 244 sheets	7.6
Mars, Incorporated	Whiskas cat food	2023	From 100 g to 85 g	15

[1] https://www.ons.gov.uk/economy/inflationandpriceindices/articles/theimpactofshrinkflationoncpihuk/howmanyofourproductsaregettingsmaller

While the phenomenon itself is fairly well recognized in the literature, in relatively few works an effort has been made to examine its impact on the measurement of inflation, particularly on the performance of a given price index. Therefore, this is one of the threads, in addition to the automation of downsizing detection, that this paper takes into consideration. Please note that our definition of the above-mentioned terms is broader than the above-cited one: we will use term *shrinkflation* when a reduction in the size of a product is accompanied by either no change in its price (narrow, existing meaning) or an apparent decrease in its price (broader meaning), which consequently means that the unit price of such a product has increased.

2 Scanner data case

Scanner data mean transaction data that specify turnover and numbers of items sold by barcodes, e.g.: GTIN, formerly known as the EAN code (International Labour Office 2004). These data are a relatively new data source for statistical agencies and the availability of electronic sales data for the calculation of the Consumer Price Index (CPI) has increased over the past 30 years. They can be obtained from a wide variety of retailers (supermarkets, home electronics, Internet shops, etc.). Scanner data have numerous advantages compared to traditional survey data collection because such data sets are much bigger and cheaper than traditional ones and they contain complete transaction information, i.e., information about prices and quantities (Eurostat 2018). Scanner data sets have a huge volume and may provide some additional information about products (such as the following attributes: their size, grammage, sale unit, color, package quantity, etc.). These attributes may be useful in aggregating items into homogeneous groups.

Nevertheless, there are numerous challenges while using these data (Białek & Roszko-Wójtowicz 2023). In addition to the usual problems associated with the procedure of scanner data (e.g.: automatic classification of products into COICOP groups, high product turnover and chain drift effect when calculating the price index, the problem of combining results obtained from different retailers and other data sources, etc.), in practice we encounter yet another reason for the bias of measuring the dynamics of scanned prices. Namely, as already mentioned in the introduction, the price index bias can also result from the downsizing phenomenon, since the actual price at which the product is sold is different from the one read from the retail chain's electronic terminal. In the case of scanner data, detection of downsizing cannot be done manually due to the huge size of this type of data. However, the literature lacks clear guidance on how to perform automatic detection of downsized products. Moreover, there are no functions and procedures available for this type of detection and

analysis in the available R packages, which are dedicated to the calculation of price indices based on scanner data. The above-presented observations were the rationale for the authors to, firstly, propose a procedure for automatic downsizing detection available in R (see Sections 3 and 4), and secondly, to verify the scale of this phenomenon for selected groups of scanner products (Section 5).

3 Proposal for an automatic procedure to detect downsizing

Our proposal for an automatic procedure to detect downsizing when working on scanner data is as follows:

- **Step 1: Matching products over time** by using all available product codes and their descriptions. In our procedure, we use the product retailer code, the EAN code, and product description including size and sales unit information. Since, as a rule, reducing the size of the product does not change the product code, but obviously affects the change of its description, it is important that the product matching is based on the text distance with an appropriately selected threshold. We used the **Jaro-Winkler text distance** measure (Jaro 1989, Winkler 1990).

- **Step 2: Extraction of size and sales unit** information using regular expressions (**regex**) based on the product description (unless the retail chain provides such information in separate columns).

- **Step 3: Price and quantity normalization** so that they relate to the unit of sale.

- **Step 4: Detection** of codes for matched products that, in the face of a **decrease in size**, have recorded an **increase in unit price**.

Please note that Statistics Poland follows steps 1-3 in the standard procedure of dealing with scanner data when counting inflation. Until now, however, although the procedure included normalization of prices and quantities, downsized products were not treated specifically and were automatically included in the analysis after transformation. The added value of the procedure in question is Step 4, which detects such cases, analyzes them statistically and places them in time. By identifying such products in detail, we are also able to take a closer look at the incorrect classifications described in Figure 4 and take further steps against them. Thus, steps 1-4 as a whole, in our opinion, are an important part of preparing the scanner data set for final determining price indices.

4 An example

In order to demonstrate how the various steps discussed in Section 3 look in R code form, a small artificial scanner data set (i.e. *dataDOWNSIZED*) that is included in the *PriceIndices* R package (Białek 2021, 2022) was used. Fig. 1 presents the form of *dataDOWNSIZED* and suggests potential downsized products. To analyze downsized products, the `shrinkflation` function was used, which is available in the latest version of the PriceIndices package (ver. 0.1.9 available at: https://cran.r-project.org/web/packages/PriceIndices/index.html). The entire procedure of downsizing detection demonstrated by using the *PriceIndices* package is in Appendix A. Figs. 2 and 3 present the results obtained by using the `shrinkflation` function, i.e.: Fig 2 shows monthly changes in product sizes and prices and Fig. 3 summarizes the scale of downsizing in the analyzed data set.

Original prices and quantities are not standardized.

Information about the size and unit of sale included in the product description.

	time	prices	quantities	codeIN	codeOUT	description	
1	2024-01-01	20	100	1	10000000000100	coffee ABC 200g	
2	2024-01-02	20	200	1	10000000000100	coffee ABC 200g	
3	2024-01-03	30	150	2	10000000000200	coffee DEF 300 g	
4	2024-01-04	35	150	2	10000000000200	coffee DEF 300 g	
5	2024-01-05	40	300	3	10000000000300	coffee GHI 2 x 400g	
6	2024-01-06	42	400	3	10000000000300	coffee GHI 2 x 400g	
7	2024-01-07	50	200	4	10000000000400	coffee JKL 250 ml	
8	2024-01-08	50	200	4	10000000000400	coffee JKL 250 ml	
9	2024-01-09	52	100	4	10000000000400	coffee JKL 250 ml	
10	2024-01-10	70	50	5	10000000000500	coffee super 0,5 l	
11	2024-01-11	50	100	7	10000000000900	cofee extra	
12	2024-02-01	18	100	1	10000000000100	coffee ABC 180g	
13	2024-02-02	18	200	1	10000000000100	coffee ABC 170g	
14	2024-02-03	31	150	2	10000000000200	coffee DEF 300 g	
15	2024-02-04	37	150	2	10000000000200	coffee DEF 300 g	
16	2024-02-05	41	300	3	10000000000300	coffee GHI 2 x 390g	
17	2024-02-06	42	400	3	10000000000300	coffee GHI 2 x 390g	
18	2024-02-07	50	200	4	10000000000400	coffee JKL 250 ml	
19	2024-02-08	50	200	4	10000000000400	coffee JKL 240 ml	
20	2024-02-09	50	100	4	10000000000400	coffee JKL 240 ml	
21	2024-02-10	69	50	5	10000000000500	coffee super 0,4 l	
22	2024-02-11	50	120	7	10000000000900	cofee extra	

Potential downsized products

Figure 1: The structure of the *dataDOWNSIZED* data set.

	prodID	grammage	unit	mean_price	size_decrease	price_increase	downsizing	description	time
1	8	0.20	kg	100.000	-	-	FALSE	coffee ABC 200g	2024-01
2	8	0.18	kg	100.000	10 %	0 %	FALSE	coffee ABC 180g	2024-02
3	8	0.17	kg	105.882	5.556 %	5.882 %	TRUE	coffee ABC 170g	2024-02
4	9	0.80	kg	51.429	-	-	FALSE	coffee GHI 2 x 400g	2024-01
5	9	0.78	kg	53.297	2.5 %	3.632 %	TRUE	coffee GHI 2 x 390g	2024-02
6	10	0.25	l	201.143	-	-	FALSE	coffee JKL 250 ml	2024-01 ; 2024-02
7	10	0.24	l	208.333	4 %	3.575 %	TRUE	coffee JKL 240 ml	2024-02
8	11	0.50	l	140.000	-	-	FALSE	coffee super 0,5 l	2024-01
9	11	0.40	l	172.500	20 %	23.214 %	TRUE	coffee super 0,4 l	2024-02

Figure 2: Monthly changes in product sizes and prices on the basis of *dataDOWNSIZED*

	stats	value
1	Downsized product shares:	----------
2	number of all products	6
3	number of downsized products	4
4	share of downsized products	66.667 %
5	turnover of all products	157400
6	turnover of downsized products	126450
7	turnover share of downsized products	80.337 %
8	Average measures:	----------
9	mean size decreases of downsized products	8.014 %
10	mean price increase of downsized products	9.076 %
11	median size decreases of downsized products	4.778 %
12	median price increase of downsized products	4.757 %
13	Volatility measures:	----------
14	standard deviation of size decreases	8.087 %
15	standard deviation of price increases	9.487 %
16	volatility coefficient of size decreases	1.009
17	volatility coefficient of price increases	1.045

Figure 3: Summary of the scale of downsizing on the basis of *dataDOWNSIZED*

5 Empirical illustration

5.1 Scanner data sets description

We used data from a big retail chain in Poland (over 510 outlets) and for the following elementary product groups:

a) food category: **yoghurt** (COICOP 5: 011441), **rice** (COICOP 5: 011111), **groats** (COICOP 5: 011123), **baked goods** (COICOP 5: 011131), **coffee** (COICOP 5: 012111);

b) non-food category: **cosmetics and hygiene products** (COICOP 5: 121321)

The considered time interval was December, 2020 - February, 2024. Data were collected and analyzed at the most disaggregated level, i.e. at the GTIN barcode level.

5.2 Price indices considered

In our study we considered the following type of price indices:

- **Unweighted bilateral indices:** the Dutot (1738) and Jevons (1865) indices;
- **Weighted bilateral indices:** the Laspeyres (1871) and Fisher (1922) indices;
- **Unweighted and weighted chain indices:** the chain Laspeyres and Fisher (CPI Manual, 2004) indices;
- **Weighted multilateral indices:** the GEKS and Geary-Khamis indices (Eurostat 2022).

5.3 Results

First of all, let us emphasize that downsizing was detected only in the case of **yoghurt** and **cosmetics and hygiene products**. Fig. 4 presents sample TRUE and FALSE detections of downsized products which were observed in the considered data sets. In the case of **rice** ('ryż' in Polish), the FALSE detection was due to an error in entering the weight of the product (a weight as high as 500 kg was entered). In the case of *groats* ("kasza" in Polish), the FALSE detection was due to a change in the multiplicity sign when specifying multipacks.

Sample TRUE detections of downsized products

	product_group	description_before_change	description_after_change
1	yoghurt	7 ZBOZ TRUSKAWKA 150G	7 ZBOZ TRUSKAWKA 140G
2	yoghurt	JOGURT TWIST 400G OWOCE LESNE	JOGURT TWIST 380G OWOCE LESNE
3	yoghurt	JOGURT NATURALNY 2% 175G LOKAL	JOGURT NATURALNY 2% 170G
4	cosmetics and hygiene products	czarne mydło 37 polnych ziół 340ml receptury zielarki	czarne mydło 37 polnych ziół 300ml receptury zielarki
5	cosmetics and hygiene products	szampon d/włosów propolisowy 400ml receptury zielarki	szampon d/włosów propolisowy 350ml receptury zielarki
6	cosmetics and hygiene products	żel p/prysz. drzewo gwajakowe 400ml yope	żel p/prysz. drzewo gwajakowe 300ml yope

Sample FALSE detections of downsized products

	product_group	description_before_change	description_after_change
1	rice	ryz brazowy 500kg britta	ryz britta brazowy 0,5kg
2	groats	kasza gryczana 4x100g	kasza gryczana 4*100g

Figure 4: Sample TRUE and FALSE detections of downsized products

yoghurt

	stats	value
1	Downsized product shares:	----------
2	number of all products	507
3	number of downsized products	16
4	share of downsized products	3.16 %
5	turnover of all products	332057271.77
6	turnover of downsized products	22483853.51
7	turnover share of downsized products	6.77 %
8	Average measures:	----------
9	mean size decreases of downsized products	4.69 %
10	mean price increase of downsized products	26.77 %
11	median size decreases of downsized products	5 %
12	median price increase of downsized products	26.92 %
13	Volatility measures:	----------
14	standard deviation of size decreases	1.68 %
15	standard deviation of price increases	15.28 %
16	volatility coefficient of size decreases	0.36
17	volatility coefficient of price increases	0.57

cosmetics and hygiene products

	stats	value
1	Downsized product shares:	----------
2	number of all products	3730
3	number of downsized products	17
4	share of downsized products	0.46 %
5	turnover of all products	560558784.56
6	turnover of downsized products	1485508.39
7	turnover share of downsized products	0.27 %
8	Average measures:	----------
9	mean size decreases of downsized products	20.06 %
10	mean price increase of downsized products	28.58 %
11	median size decreases of downsized products	12.5 %
12	median price increase of downsized products	18.41 %
13	Volatility measures:	----------
14	standard deviation of size decreases	15.99 %
15	standard deviation of price increases	34.47 %
16	volatility coefficient of size decreases	0.8
17	volatility coefficient of price increases	1.21

Figure 5: Summary of downsized products with basic stats for **yoghurt** and **cosmetics and hygiene products**

Fig. 5 presents a summary of downsized products with corresponding basic stats for **yoghurt** and **cosmetics and hygiene products**. As one can see, the downsizing phenomenon is more pronounced in the case of **yoghurt**, where the share of sales of downsized products has exceeded 6.7%. The mean size decrease was substantially bigger in the case of **cosmetics and hygiene products** (20.06% for cosmetics and hygiene products vs 4.69% for yoghurt) but the mean price increase was comparable in both data sets.

A natural question arises about the impact of downsized products on the price index score. To answer this question, the price indices discussed in Section 5.2 were determined with estimates made in two variants: on the entire data set and on the set reduced by downsized products. Fig. 6 – 9 present a comparison of all discussed price indices with respect to these two variants. As could be predicted, the impact of downsized products on the value of the price index is clearly greater in the case of **yoghurt**, since simply its share of sales is clearly bigger here (with a comparable average price change accompanying downsizing).

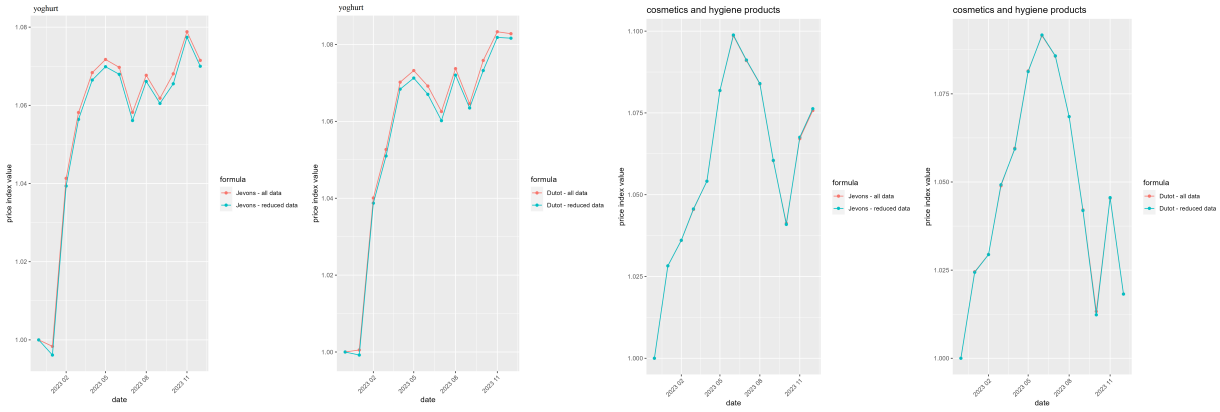

Figure 6: Comparison of selected unweighted price indices with respect to the entire and reduced data sets

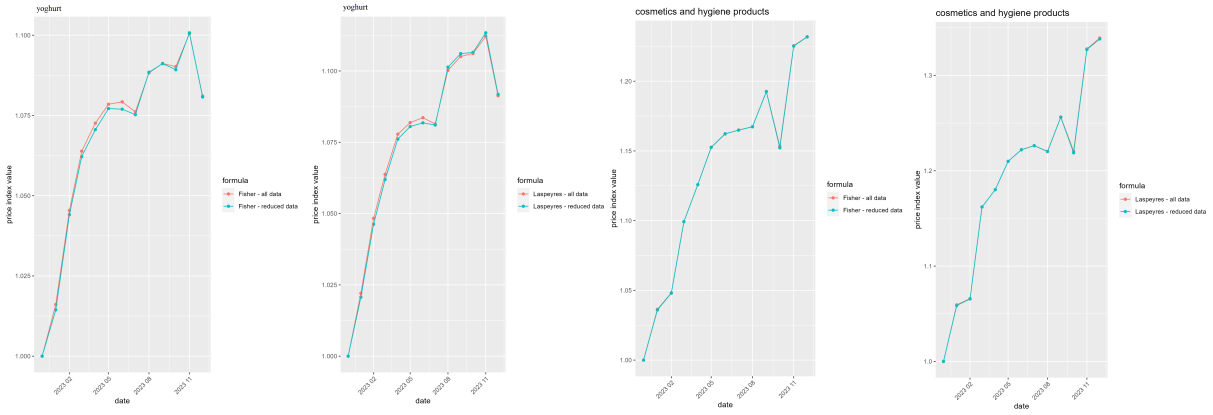

Figure 7: Comparison of selected weighted price indices with respect to the entire and reduced data sets

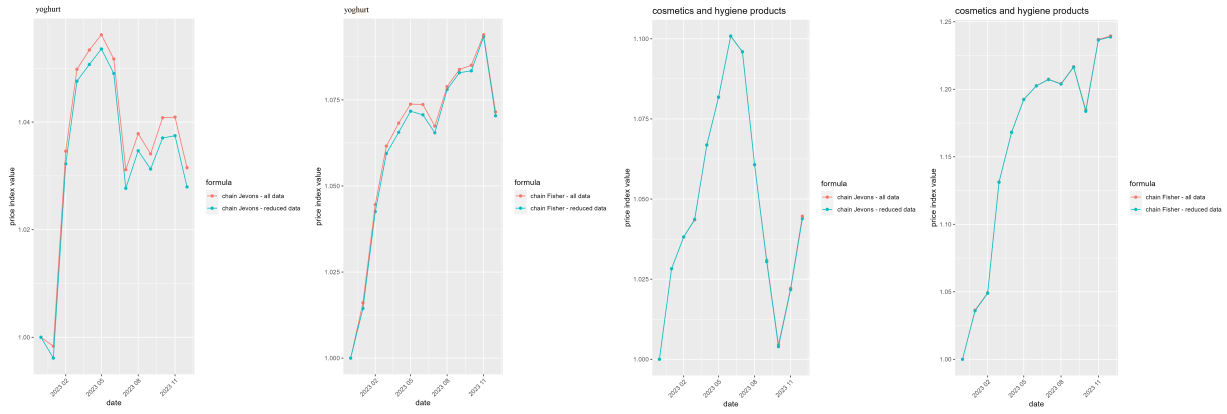

Figure 8: Comparison of selected chain price indices with respect to the entire and reduced data sets

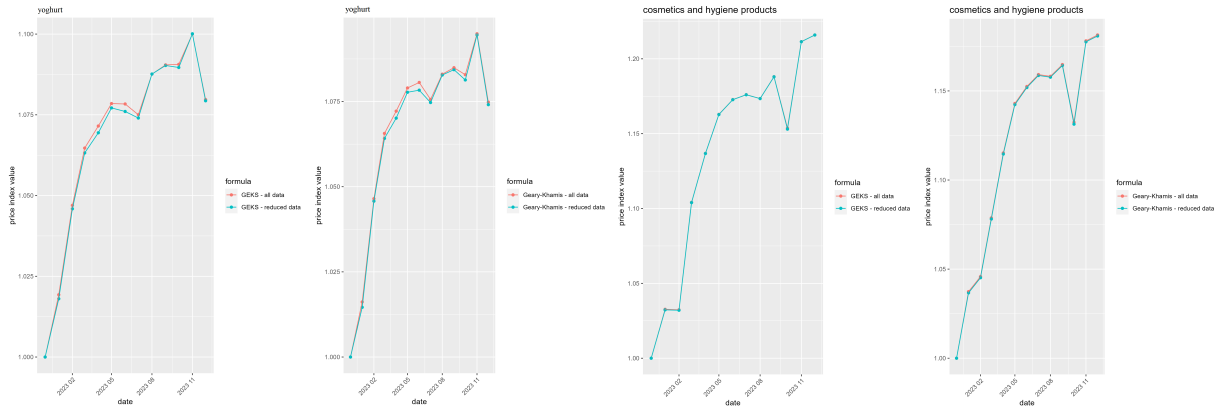

Figure 9: Comparison of selected multilateral price indices with respect to the entire and reduced data sets

6 Conclusions

- The automatic downsizing detection procedure can be effective as long as correct product matching based on both product codes and an appropriate measure of text distance is carried out. The procedure also requires normalization of prices, so information on the product's weight and selling unit is needed.
- When the share of sales of downsized products is small (less than one percent) then their impact on the value of the price index will be marginal. This observation holds true in the context of cosmetics and hygiene products. In the case of a few percent (or more) of sales of downsized products, we may see a measurable overestimation of the price index. We observed this phenomenon in the case of **yoghurt**, where this overestimation could exceed 0.35 p.p. (please note that this is only one elementary group). Chain indices seem to be the most sensitive to downsized products, while unweighted bilateral ones seem to be the least sensitive.

Appendix A R script providing figures in Section 4

library('PriceIndices')

1. **Data matching** over time:

   df <- **data_matching**(data = dataDOWNSIZED,

   start="2024-01", end="2024-02",

   codeIN = TRUE, codeOUT = TRUE, description = TRUE,

   onlydescription = FALSE, precision = 0.9, interval = FALSE)

2. **Extraction** of information about product grammage:
   df <- **data_unit**(df, units = c("g|ml|kg|l"), multiplication = "x")

3. **Price and quantity normalization**:
   df<-**data_norm**(df, rules = list(c("ml","l",1000), c("g","kg",1000)))

4. **Downsized products detection**:
   result <- **shrinkflation**(data = df,
   start = "2024-01",end = "2024-02",
   prec = 3, interval = FALSE)
   result$changes
   result$products_downsized
   result$df_downsized
   result$df_reduced
   result$df_summary

References

Białek, J. (2021), 'PriceIndices – a new R package for bilateral and multilateral price index calculations', *Statistika – Statistics and Economy Journal* **36**(2), 122–141.

Białek, J. (2022), 'Scanner data processing in a newest version of the PriceIndices package', *Statistical Journal of the IAOS* **4**(38), 1369–1397.

Białek, J. & Roszko-Wójtowicz, E. (2023), 'Potential reasons for CPI chain drift bias while using electronic transaction data', *Technological and Economic Development of Economy* **29(2)**, 564–590.

Domitrovic, B. (2009), *Econoclasts : the rebels who sparked the supply-side revolution and restored American prosperity*, Wilmington, Del. : ISI Books.

Dutot, F. (1738), *Reflexions Politiques sur les Finances et le Commerce*, The Hague: Les Freres.

Eurostat (2018), 'Practical guide for processing supermarket scanner data', Harmonised Index of Consumer Prices.

Eurostat (2022), *Guide on Multilateral Methods in the Harmonised Index of Consumer Prices*, Luxembourg: Publications Office of the European Union.

Fisher, I. (1922), *The making of index numbers: a study of their varieties, tests, and reliability*, Vol. xxxi, Houghton Mifflin.

Giner, D. (2022), 'Hi shrinkflation, meet your cousin shitflation', *THE BLIND SPOT, https://the-blindspot.com/hi-shrinkflation-meet-your-cousin-shitflation/* .

International Labour Office (2004), 'Consumer Price Index manual: Theory and practice.', Geneva.

Jaro, M. (1989), 'Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida', *Journal of the American Statistical Association* **84**(406), 414–420.

Jevons, W. S. (1865), 'On the variation of prices and the value of the currency since 1782', *Journal of the Statistical Society of London* **28**(2), 294–320.

Laspeyres, K. (1871), 'Ix. die berechnung einer mittleren waarenpreissteigerung.', *Jahrbücher für Nationalökonomie und Statistik* **16**(1), 296–318.

Winkler, W. (1990), String comparator metrics and enhanced decision rules in the fellegi-sunter model of record linkage, *in* 'Proceedings of the Section on Survey Research Methods. American Statistical Association', p. 354–359.