

Coverage Bias and the Effect of Re-Launches in Scanner Data: A Coffee Index

Åsa Bilius, Olivia Ståhl and Can Tongur
 firstname.lastname@scb.se
 Statistics Sweden

Meeting of the
 Group of Experts on
 Consumer Price Indices
 7th to 9th of May, 2018
 Geneva

Background

Scanner data for daily necessities was implemented in the Swedish CPI in 2012. A fixed basket approach is used in which a probability sample of items (GTIN:s) is selected for the base period December $t-1$, and prices then surveyed throughout the year. Every month the sample undergoes a manual examination where items exhibiting reduced sales are replaced. When needed, quantity adjustments are made to render the replacements comparable.

This study compares the current approach to alternative and competing scanner data methods. The product group Coffee exhibited a lot of dynamics during year 2017 and has been the subject for this analysis. Several manual replacements were made due to simultaneous changes in package size and GTIN, exemplifying the relaunch problem.

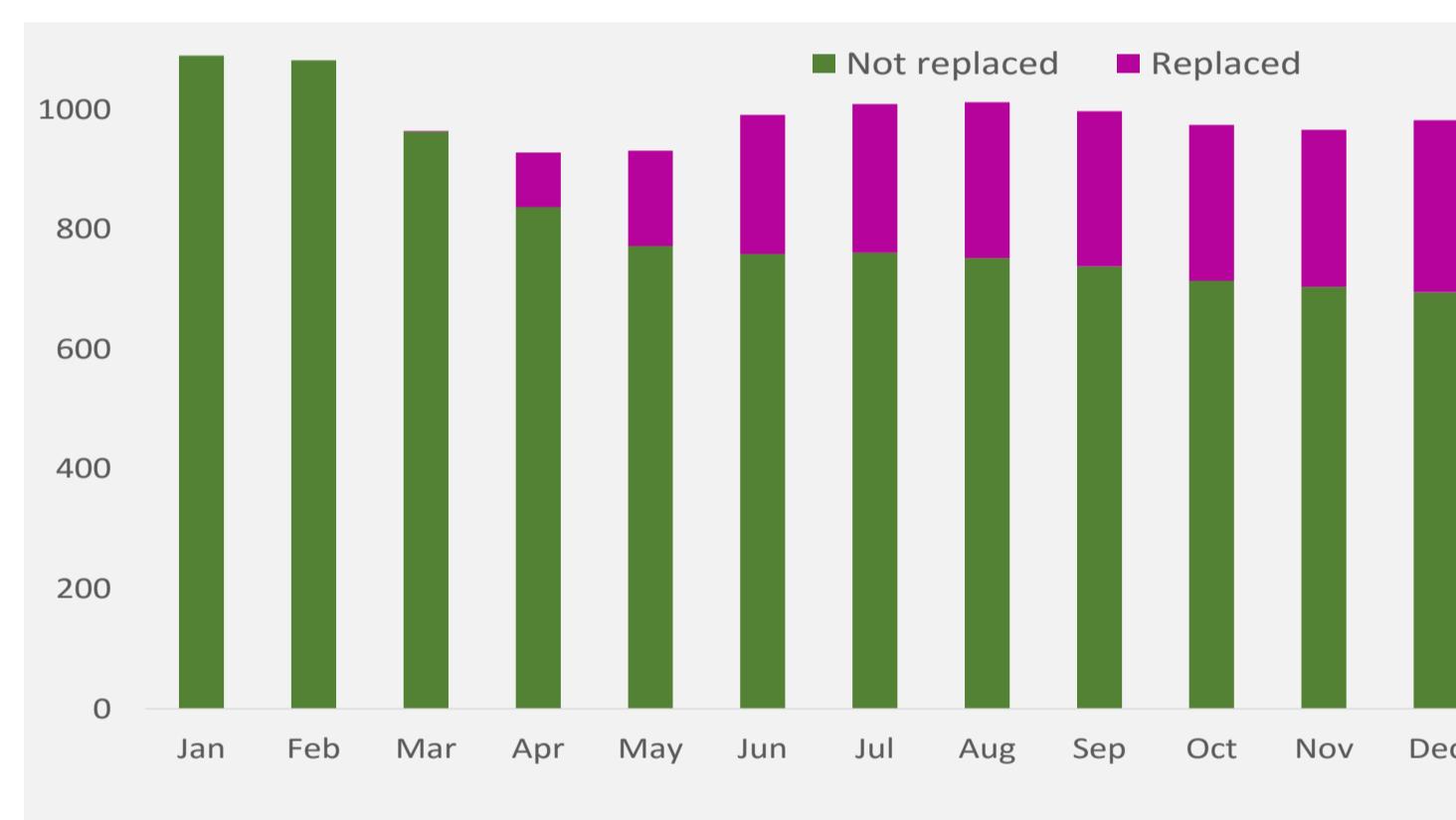


Figure: Number of manual replacements made during 2017, for the product group coffee.

With the current approach, relaunches are handled only when replacements are made. Completely new products are not included in the index. The coverage proportion, as compared to using the full data set, therefore risks to decrease throughout the year.

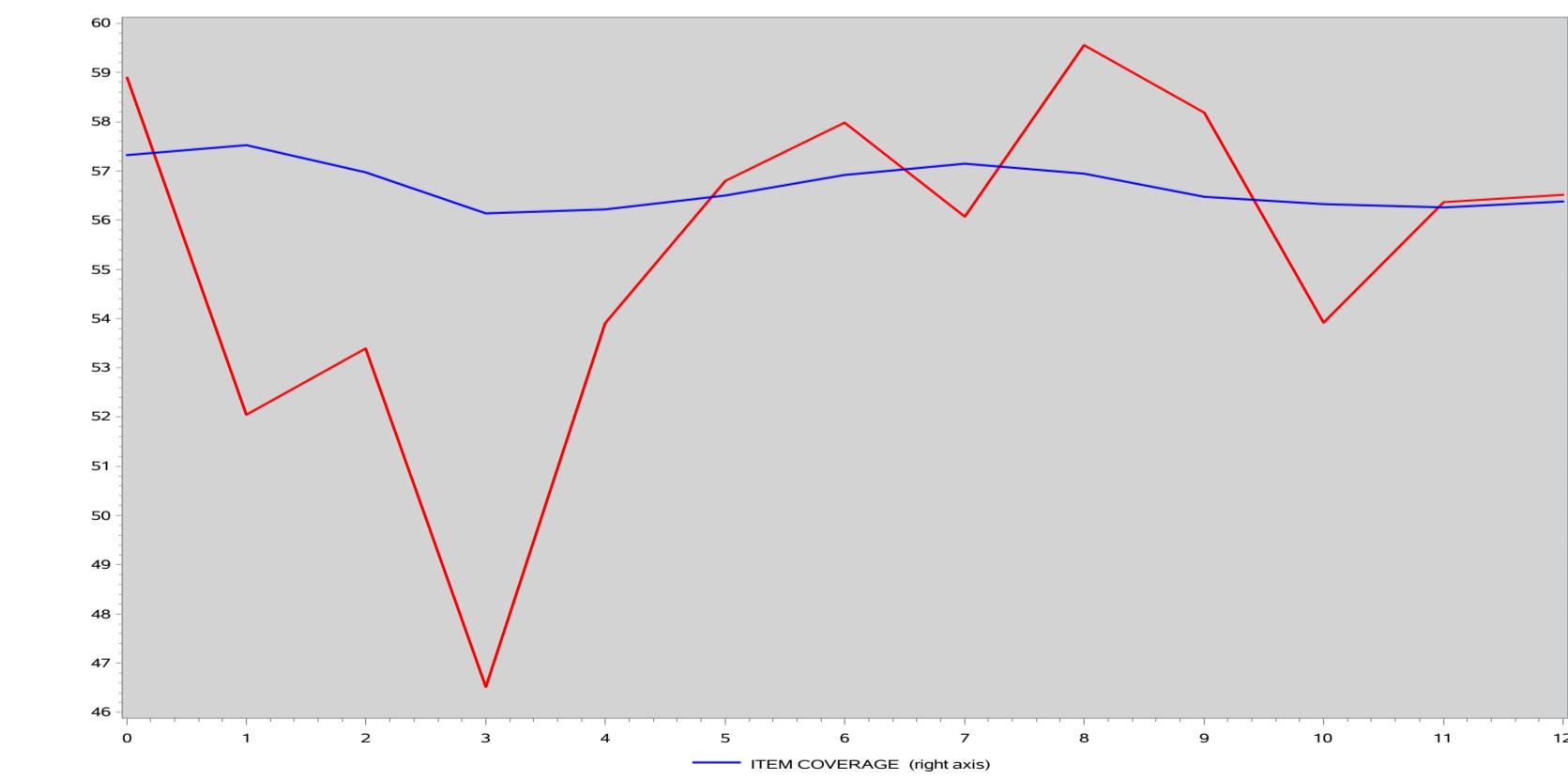


Figure: Coverage proportion for the current approach in terms of the number of product offers and (unweighted) total sales.



Fixed Basket Indices

Direct comparisons of month m with base period ($0 = \text{December } t-1$).

- **Current method (sample data)**

- A π_{ps} sample of approximately 1100 product offers.
- Weighted geometric index, with weights taking into account of the π_{ps} sampling design.

- **Unweighted fixed basket method - extended data**

- All of the data is used.
- Approximately 5800 product offers.
- Unweighted geometric index.

- **Weighted fixed basket method - extended data**

- All of the data is used.
- Approximately 5800 product offers.
- Weighted geometric index, with weights equal to yearly sales.



Figure: Fixed basket indices for coffee prices in 2017.

The number of unique GTIN:s surviving from the base period in the extended data set is decreasing throughout the year.

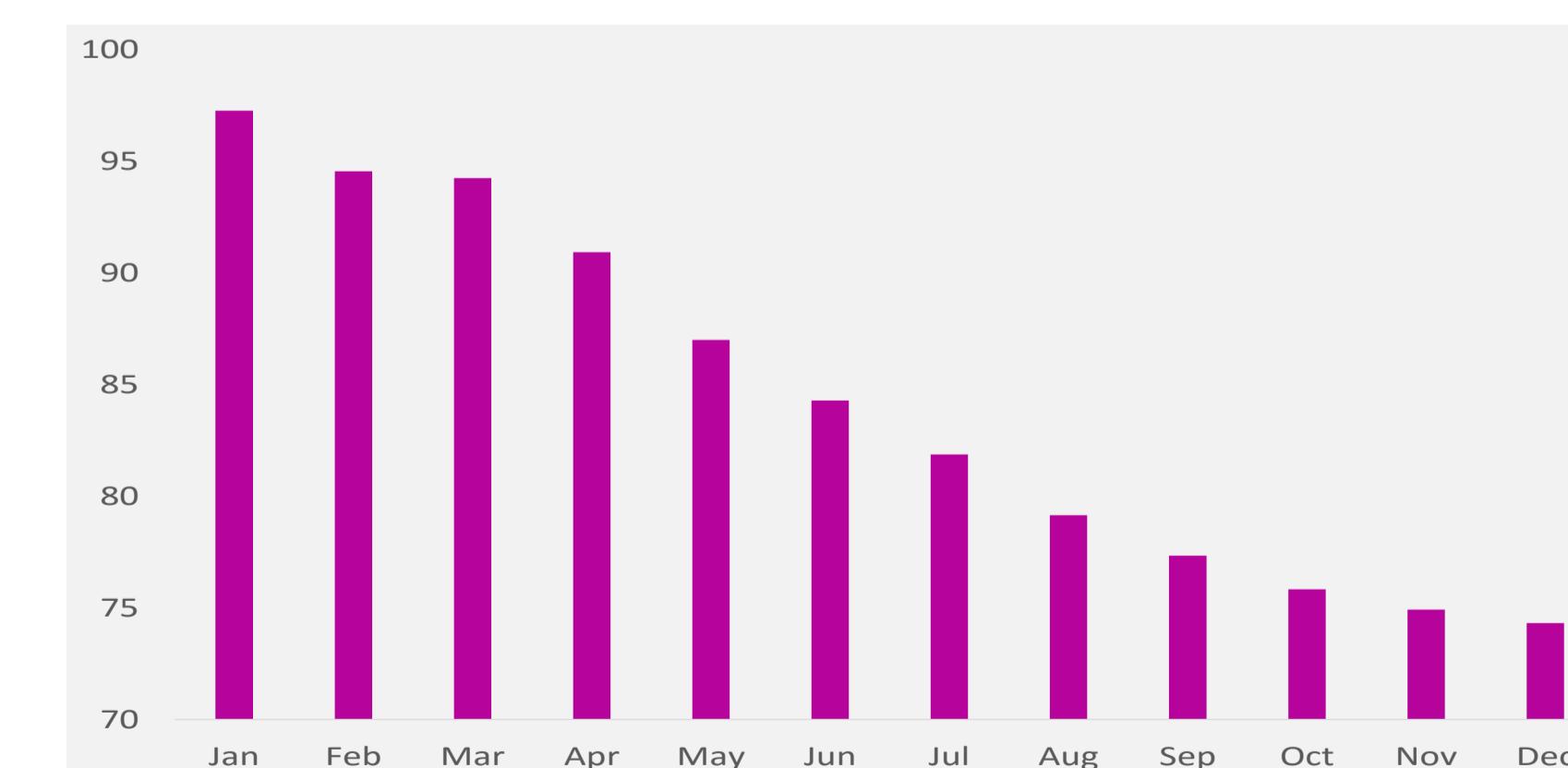


Figure: Unique GTIN:s each month, as a proportion of the number of unique GTIN:s available in the base period for the extended data set.

Product dynamics

The dynamics of the product can be described in terms of flow, inflow and outflow. Here we define flow as the number of items sold during a particular month, which were also sold during the base month. Inflow is defined as the number of items sold that were not sold in the base month, and outflow as the number of disappearing items. In the figure below, flow, inflow and outflow frequencies are presented as percentages with respect to the total number of items in either the base month or the month under study.

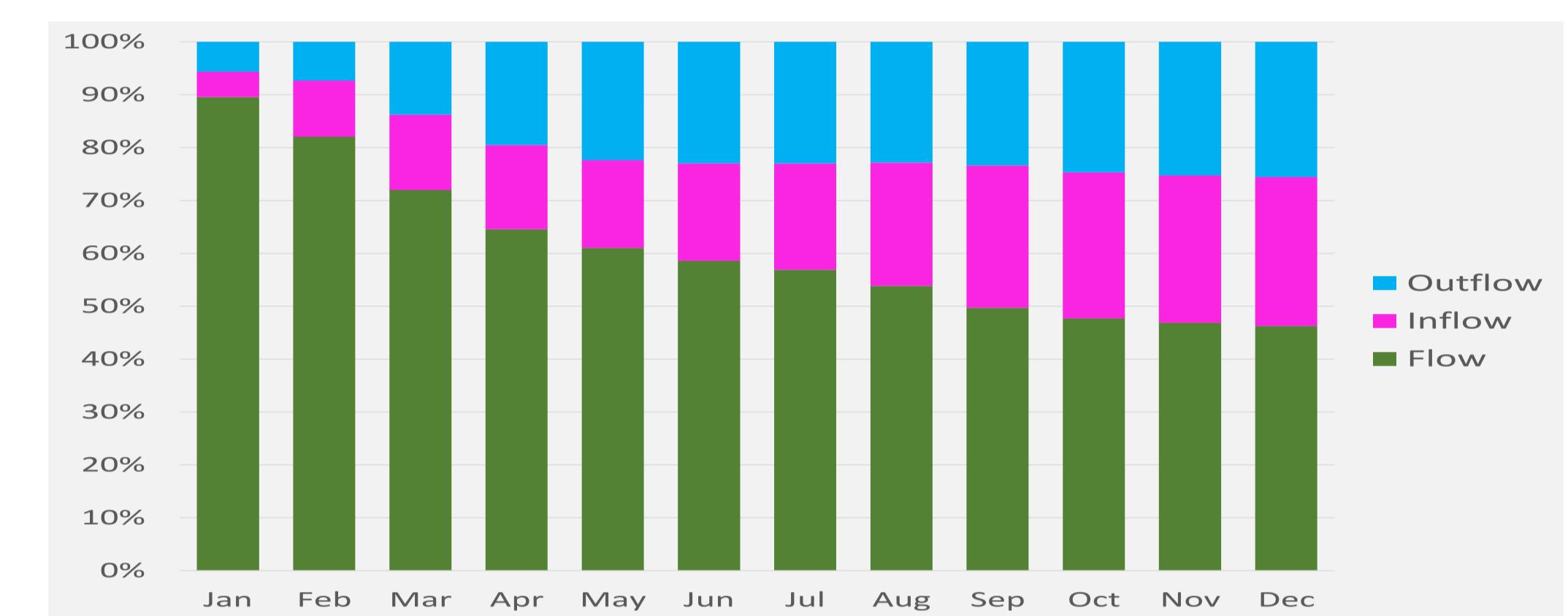


Figure: Flow, inflow and outflow in the coffee scanner data.

Matched Items Chained Indices

Indices based on monthly matched items' price ratios.

- **Unweighted geometric index**

Chained Jevons index (no cutoff)

- **Unweighted geometric index with cutoff**

Chained Jevons index using only items which account for at least 1% of total sales during the two consecutive months $m-1$ and m

- **Weighted geometric index (fixed weights)**

Yearly weights consisting of total sales in 2017 (no cutoff)

- **Weighted geometric index (monthly weights)**

Chained Törnqvist index (no cutoff)

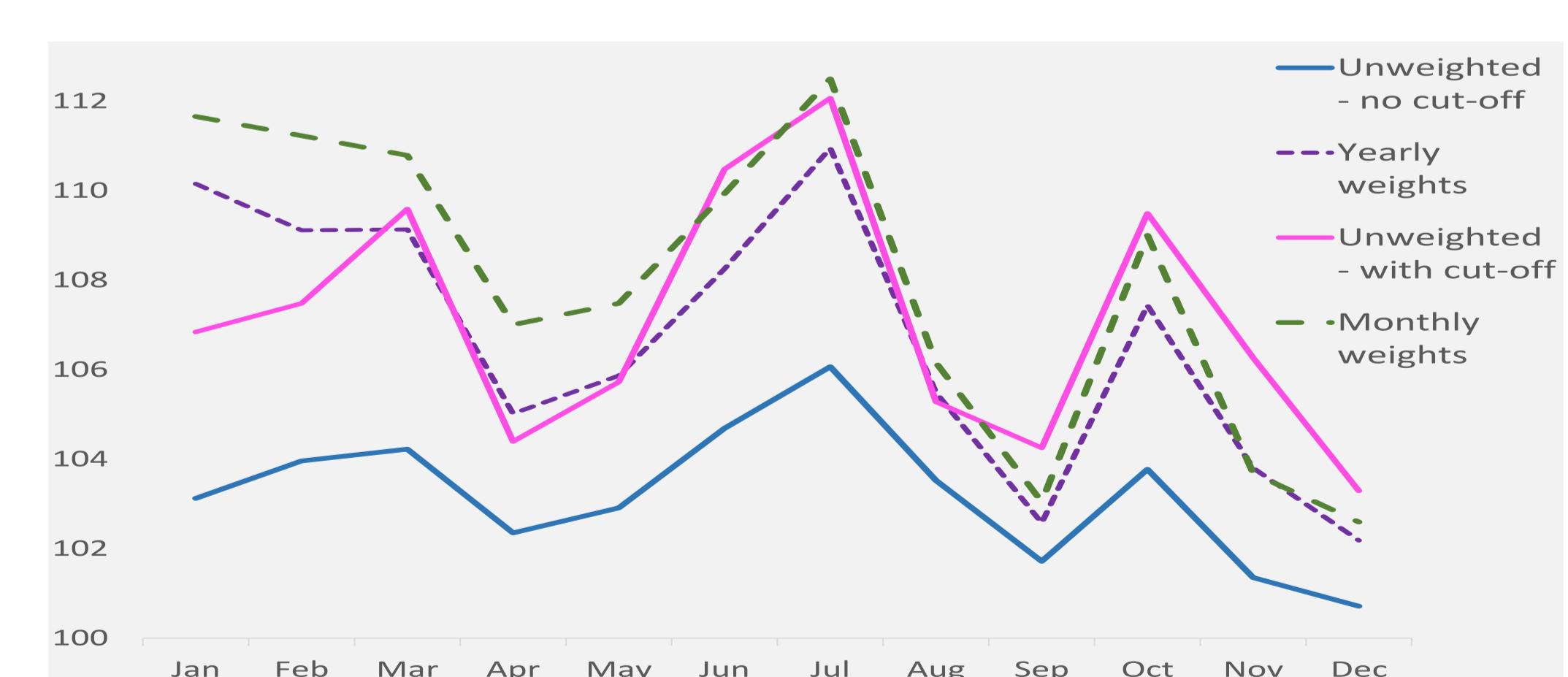


Figure: Monthly chained matched items indices.

Geary-Khamis Index

Geary-Khamis index is applied using data from the full year to estimate the quality adjustment factors (i.e. not a real-time index).

The two-stage approach is proposed in recent literature as a way of solving the relaunch problem that is otherwise bypassed. It embodies grouping of items, where homogeneity is sought within each group.

Groupings for the GK-method

Three dimensions are used to construct the groups: size, brand and type of coffee (e.g. Presso, Grounded, etc.). The size dimension divides coffee packages into 5 categories: 0-99 grams, 100-399 grams, 400-749 grams, >750 grams and a separate category for piece-packages (Capsule/Pod coffee). For the brand dimension all unique brands are specified one-by-one. The type of coffee dimension is a basic categorization of purpose/sort. The first type categorization is universal over all chains (crude groupings) whereas the second type is a retailer-specific categorization. These may or may not coincide, depending on the level of detail in retailer groupings.

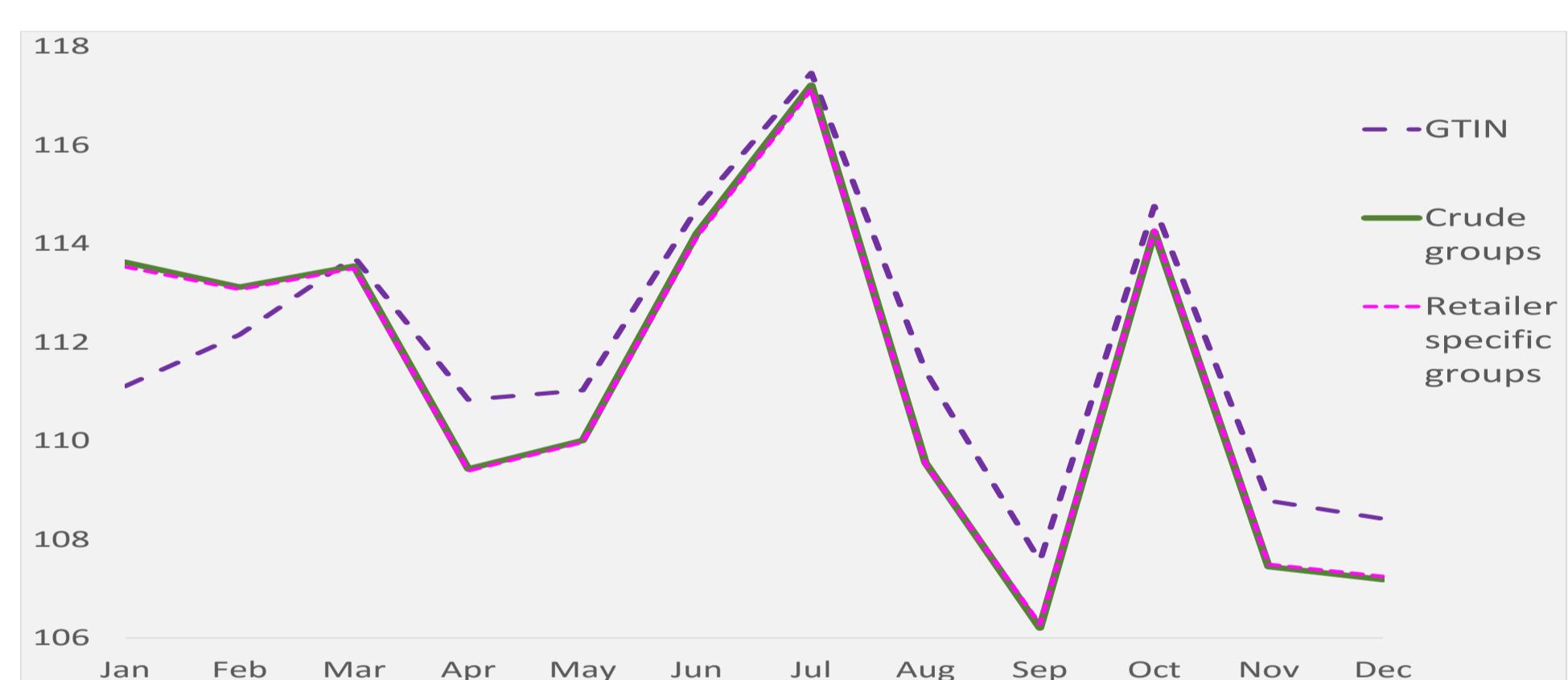


Figure: Geary-Khamis Indices based on GTIN and on homogeneous groups.

When forming homogeneous groups, variation in the number of GTIN:s appears not to systematically affect index.

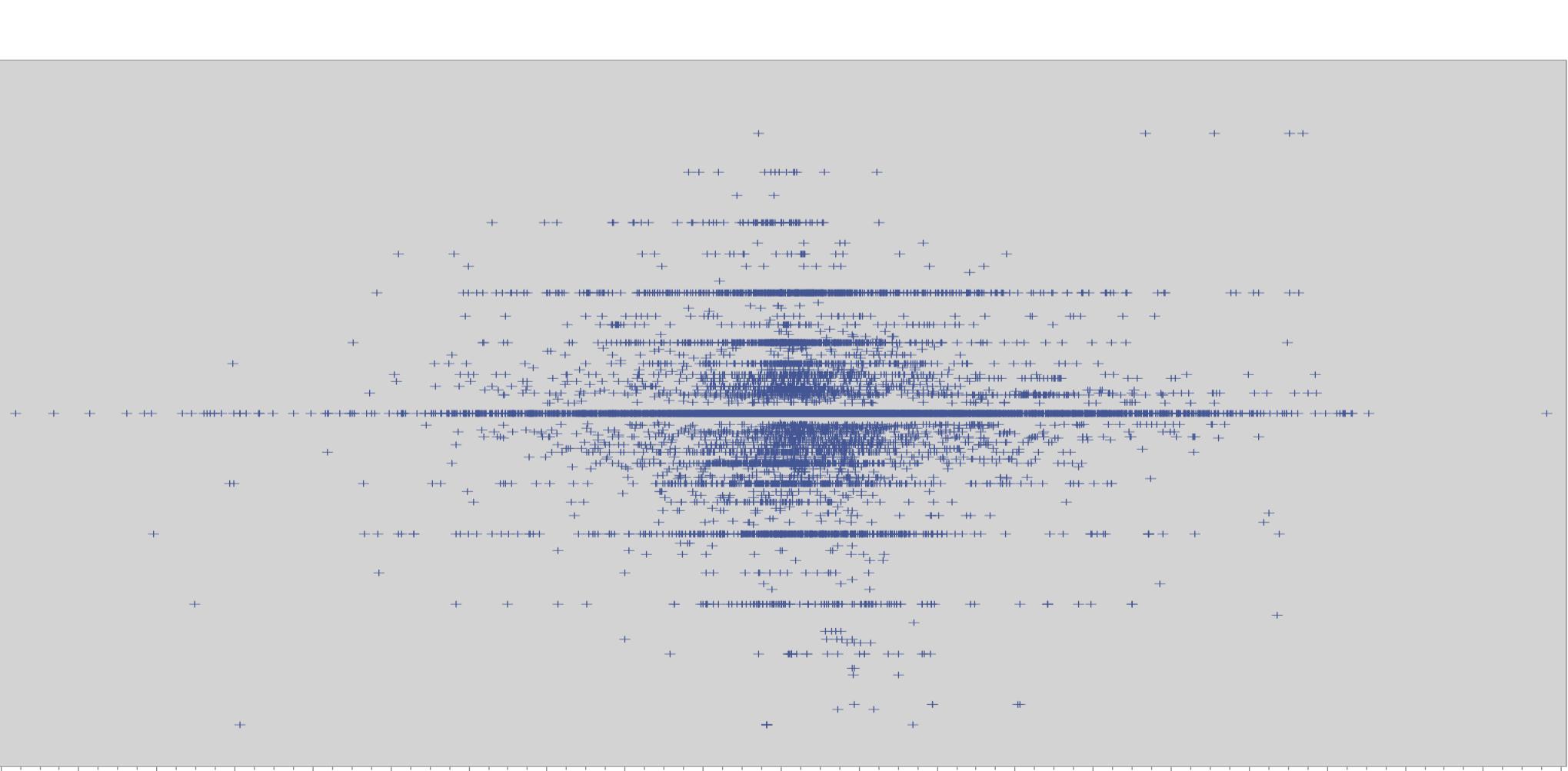


Figure: Group content change and effects on index, for retailer specific grouping.
 (Y-axis: logarithm of the ratio of the number of GTIN codes in a particular group as compared to the base month. X-axis: logarithm of price ratios, compared to the base month.)