How did Statistics Iceland start using scanner data?

Heiðrún Erika Guðmundsdóttir, Statistics Iceland, heidrun.gudmundsdottir@statice.is

Statistics Iceland, Borgartúni 21a, IS - Reykjavík 150, Iceland
Tel: +354 528 1200, Fax: +354 528 1299

The implementation

In April 2016 Statistics Iceland started using prices obtained from electronic cash register data (scanner data) from three data providers in its CPI and HICP. The delivery of data from the first data provider started about a year earlier but the other two followed shortly. The year was used to study the data and prepare for their implementation.

Scanner data is used in COICOP	
01	Food
02.1.3	Low/non-alcoholic beer
05.6.1	Non-durable household goods
09.3.4	Pet foods
11.1.1	Ready-made meals and sandwiches
12.1.3	Articles and non-electric appliances for personal care

The timing of the implementation was chosen in April as that is the month when the CPI is rebased annually. It would have been desirable to have a longer period than a year for thoroughly testing the data and building infrastructure, but the pressure to advance into using the data was considerable, yet based on expectations of important quality improvements. Due to the time constraint the change was limited to a change in data source and thereby gaining some of the expected benefits early, while continuing to work on further progress. Among the benefits of this action was having more extensive data, both temporal and spatial, and better knowledge of the actual sales volumes of groceries. By collecting data daily, the likelihood of missing prices is diminished, and so is the recovery time if the data delivery fails. The latter is important, especially this early in the process when delivery processes are still being streamlined. In due course the focus will move towards better utilisation of the data.

By starting using scanner data, an important milestone has been reached, but many more are still ahead. These include increasing utilisation of the data set, strengthening the work processes, training of staff, studying the data further and increasing automation in coding. By taking this stepwise approach some of the instant benefits of utilising scanner data have been accomplished, while acquiring important knowledge and experience for further implementations.

A new data source only

Implementing scanner data only as a new data source means that the method of aggregation of indices remains unchanged. The scanner data are entered into the aggregation processes by the same means as the manually collected data were. The gradual aggregation of prices in five steps (see figure on flip-side) is equally valid for both manually collected data and scanner data, however on closer look it can be seen that steps I and II are mere idle steps for manually collected prices unless the price collector would have returned measurements from more than one visit per outlet per month. On the other hand the method becomes fully utilised as scanner data is received on a daily basis.

Statistics Iceland

As the focus was on just changing the data source many of the possibilities of using the scanner data are yet to be exploited. The bulk of data received has not been linked with COICOP. The sample was chosen manually with reference to a list of the most sold items for each retailer within the retailers own classification of previously sampled products. One future project is the mapping to COICOP in order to use more automated and dynamic approaches for sampling. Another future aspiration is to omit step I in the diagram below and use a week average price instead of an average of daily average prices. This change would be in accordance with Eurostat's recommendations on processing scanner data. Implementing these changes will have to be announced in advance as major methodological changes.

Method of aggregation
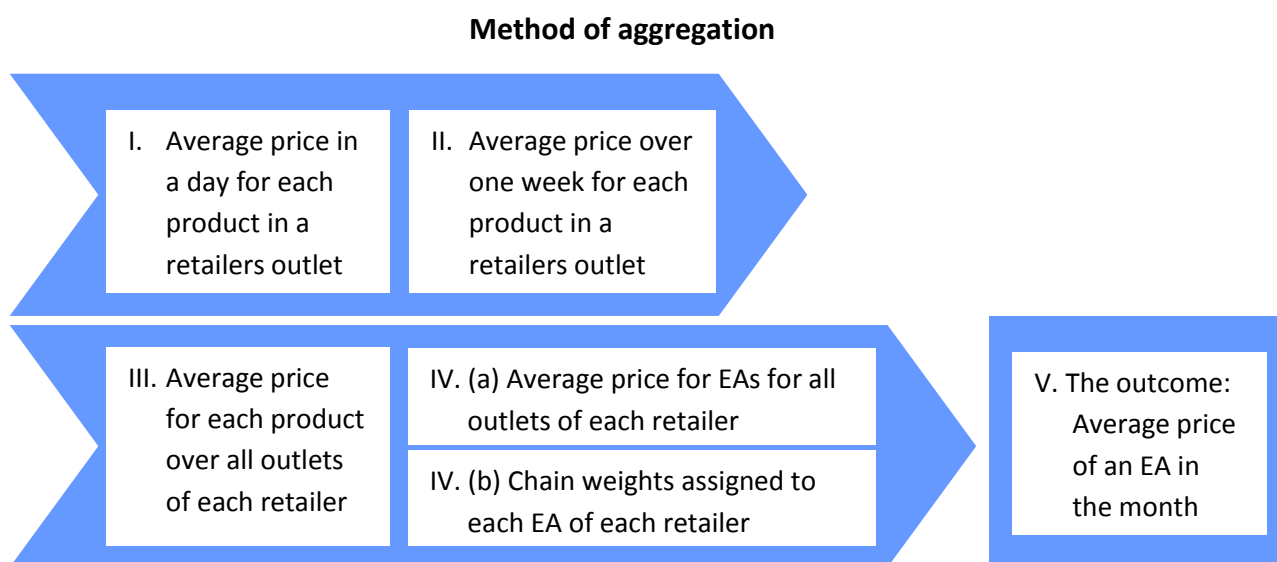
A five step method of aggregation, where steps I to IV describe unweighted geometric means. One week refers to the 7 days of price collection in the middle of the month as stated in the Icelandic law on the consumer price index. The average price during this week is the price of the month. Chain weights are deducted from the Household Expenditure Survey per EAs. The compilation method is a fixed basked, weighted Jevon index.

Comparisons of prices from scanner data and manual price collection during the implementing period were successful. One of the complications faced when implementing scanner data was the choice of the unique identifier to be used for linking prices between months. At first a very strict combination of barcode, retailer specific product number and classification was used but turned out to be too strict. At the moment a system of CPI_IDs issued manually over a range of barcodes are being used. More data analysis is needed to determine if retailer specific item codes can be used as the unique identifier. The new CPI_IDs are in some cases more detailed than the product descriptions used in manual price collection. This, along with better temporal coverage during the price collection period has resulted in less fluctuation due to product variances and missing prices.

Chain weights

The use of chain weights was the keystone to implementing the use of scanner data as it was done, as it easily enabled mixing manually collected prices and scanner data prices. Even though scanner data was retrieved from the biggest retail chains in the country it was decided to continue manual price collection from some small retailers, at least while scanner data as a main data source and the newly initiated data collection processes had proved themselves to be reliable over a longer period. Should the new processes have failed the old processes would still be accessible and functional.

Statistics Iceland