

## Scanner Data: Advances in Methodology and New Challenges for Computing Consumer Price Indices\*

Marie Leclair, Isabelle Léonard, Guillaume Rateau, Patrick Sillard, Gaëtan Varlet and Pierre Vernédal

### Key Question

Official statistics are usually based on survey or administrative data. Access to private data offers new possibilities while raising new statistical processing issues. In the field of price statistics, scanner data, that is, data collected at the time of the consumer's checkout, have long been used by supermarket chains. The article shows how they can be used to calculate the French consumer price index.

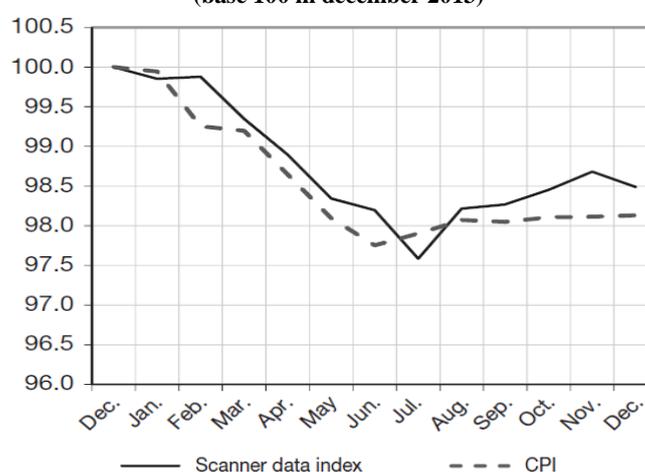
### Methodology

The article uses scanner data, collected on an experimental basis, to implement the solutions considered to exploit them (index aggregation formula, adjustment of quality effects, etc.). When different methods are possible, it offers variants. Finally, it compares the indices obtained using scanner data with the CPI as currently calculated, i.e. based on the prices manually collected by Insee surveyors.

### Main Results

- Scanner data allow more detailed and accurate statistics to be produced because they are exhaustive in their field.
- The aggregation of the prices recorded to calculate an index is modified by the availability of new information on the quantities purchased for each product as well as by the much larger number of records (continuously over the month, all products purchased).
- The replacement of a product in the basket can be automated by making a quality adjustment; the latter benefits from retrospective knowledge, in the scanner data, of the past prices of the replacement products.
- The classification of products, the identification of commercial relaunches that were previously done manually, can be automated by using a barcode referential.
- Overall, the consumer price indices calculated on the basis of scanner data are close to those currently calculated on the basis of manual collection by surveyors in the points of sale.

Consumer price indices for two item headings and indices calculated solely using scanner data, 2014  
(base 100 in december 2013)



Coverage: For the CPI, all forms of sale; for scanner data, super and hypermarkets; scanner data exclude promotional data. Sources: CPI, Scanner data from four retailers with a combined 30% market share.

### Message

The use of scanner data is an opportunity for the calculation of consumer price indices. Although the volume and nature of the data require the use of new statistical and IT solutions (e. g. new aggregation methods, big data technologies), the automation of a number of processing operations and the use of external referentials that enrich the data make it possible to preserve the current conceptual and methodological framework of consumer price indices while fully exploiting the new features of scanner data (all the data, knowledge on quantities sold).