

# 15<sup>th</sup> Meeting of the Ottawa Group

## 10 – 12 May 2017

### Session 9: Implementing scanner data

#### *Implementation of the treatment of the scanner data in France*

**Guillaume Rateau, INSEE**

The French statistical institute plans to modernize its consumer price index by compiling the scanner data of the large and middle-size retailers. To this end, the institute has available very itemized data. On the one hand, they are detailed by GTIN, day and outlet. On the other hand, each GTIN is described, in a structured way, by almost all the characteristics that appear on the product or on its packaging: brand, weight, volume, flavour, label... These characteristics are updated from photos of the products within a very short period of time after their appearance or their change. These data amount to circa 50 millions observations a day, since beginning 2013. The aim is to compute soon scanner data indices, in parallel of the usual price index production.

Since the beginning of the project, several strong methodological choices have been made and they frame the last methodological studies needed for the implementation. Firstly, the institute chose to keep the usual price index concepts. Hence scanner data are compiled through the use of a fixed yearly bundle of GTIN x outlet called products. Secondly, following a study<sup>1</sup> of the ways consumer can substitute their purchase of goods, elementary aggregates are computed by outlet, consumption segment and *a priori* day of the month.

These choices leave unclear several methodological questions: Should the elementary aggregates be computed by day, by week or by month? How to choose, at the level of outlet x consumption segment, products to be put in the basket? Should we put all of them? When computing the elementary aggregates, should we use annual weights by products and by day of the month or of the week? When replacing products, how to choose the replacing products? As products may not be sold every day in each outlet, how to impute missing prices?

This paper will aim at studying the main points answering these questions.

---

<sup>1</sup> Varlet G (2015) Utilisation des données de caisse de la grande distribution pour le calcul de l'indice des prix à la consommation.