Calculation methodology for the *scanner data* in the CPI and HICP

Sub-directorate General for Family Budgets and Prices
February 2020
1. Introduction

Most of the prices used in the calculation of the Consumer Price Index (CPI) are collected through personal visits to the establishments. This system guarantees that the price collected is the price of sale to the public, and that the product for which the price is collected remains the same over time. In both cases it is the interviewer who directly certifies compliance.

The CPI sample is designed so that it represents the majority of the population, and covers the most relevant aspects of their behaviour as consumers. Thus, the criteria used for the selection of municipalities is based on the population size and geographic location within each province. In turn, in each of the municipalities, the most representative commercial areas are defined, in which the establishments with the largest influx of public are chosen.

On the other hand, for the determination of the number of prices to be collected for each article, their variability is taken into account (the higher the frequency of variation, the more prices are needed) and an article’s importance or weighting (the higher an article’s relevance in the shopping cart, the more necessary it is to devote more resources to ensure the measurement's accuracy).

Thus, in the 2016 Consumer Price Index (CPI), the number of prices that are processed monthly is approximately 220,000. Most of them are collected in the approximately 33,000 establishments included in the sample, distributed among the 177 selected municipalities.

The price collection system is different depending on the article's characteristics. Thus, items whose prices are the same throughout the country or their management are optimized if done centrally, and are collected in the central services of the INE. However, for most items in the shopping cart, price collection is done by visiting the INE staff at the establishments of the selected sample.

This sample selection system and the information collection procedures guarantee an indicator with a high degree of quality and precision, at the level of the best CPIs in the surrounding countries. However, the INE works constantly to develop new methods that allow it to produce its statistics more efficiently, reducing collection costs, increasing the accuracy of its estimates and, above all, trying to reduce the burden of response on informants.

As a result, in recent years the INE has in many of its statistics implemented the use of new methods and techniques for obtaining information, based on the exploitation of administrative records and the use of electronic collection devices.

In the case of the CPI, the collection of prices in establishments through electronic devices will be a reality in the short term. Regarding the obtaining of information from company records, the INE has worked in recent years on the development of a method called scanner data, which consists of using company databases, with information on sales of their products, as substitutes for the collection of prices in establishments.
This document presents the main characteristics of this method, and details the procedure for calculating indices based on the information obtained from the databases, as well as their integration into the CPI.

2. What is scanner data?

Scanner data is the process of obtaining and using information from company databases for calculation of the CPI. These databases contain the record of sales of each product made in the check-out line of every establishment. Usually, this information consists of the number of units sold and the income for each of the products sold, coded according to each company's classification.

The scanner data method is already being used in several countries around us, as it is a more efficient, accurate and complete alternative to measure price developments. The European statistical office, EUROSTAT, thus promotes its use in the area of price index harmonization for EU member states. Spain has, of course, joined the proposal, and INE has started a pilot project with the objective of evaluating all aspects of the possible implementation of scanner data in the calculation of the CPI and the HICP.

Over the past three years, the INE, based on the experience of other countries, has developed the most appropriate methodological model for the treatment of information from supermarket and hypermarket chains and their possible integration into the calculation of the CPI.

Once the methodological design phase has been completed, the implementation process has begun, whose starting point will be January 2020 and which will be carried out gradually as information from the different chains of establishments becomes available.

In this first phase, the information to be incorporated into the CPI will be related to consumer products (packaged food, beverages, household cleaning and maintenance products, parapharmacy products, pet food and pet products and personal care items).

The immediate consequence of the implantation of scanner data is the elimination of the collection of prices in the establishments of the chains that contribute their databases. But this is not the only change. Due to the nature of the information contained in the databases (fundamentally, number of units sold per product and income obtained) the calculation procedure cannot be the same as that used up until now in the CPI.

Since its inception, two major issues had to be taken into account to address the adaptation of this new system of obtaining information to the usual CPI methods:

A. Obtaining of Information

One of the fundamental reasons why scanner data is a method on the rise in most of the surrounding countries is that the information necessary to carry it out does not require specific elaboration by the reporting companies.

The basic information for the development of the project is, for each product coded by the company and in each establishment, the following: income or unit values, quantities,
product name, description (if there is any field where it is distinguished), code (EAN ¹ or the intern that the company uses) and internal classification.

The format of the databases, the system of transmission of the same and the other aspects related to the disposition of the information must be decided by the company, so that the load is the smallest possible. Obviously, this conditions the work that the INE must address, since processes must be designed to adapt to the different contents, formats and access systems of the databases of each chain that participates in the project. Therefore, there will be as many procedures for obtaining and managing databases as reporting companies.

B. Information Use

One of the main dilemmas presented with using this type of information in the CPI is of a conceptual type: the incorporation of data on company sales requires changes in the methods and definitions traditionally used in this indicator’s calculation methodology. The main conceptual challenges are the following:

- **Differences between the item price and unit value**

The CPI measures, by definition, the evolution of the prices of goods and services acquired by households. Therefore, the sale price to the public in each establishment is collected. The use of scanner data, however, changes this philosophy since it requires that for each EAN code, its unit value (total revenue/total units sold) be used, but not the price itself.

In reality, the unit value does not correspond to a single real transaction but represents all those carried out over a fixed period of time. This implies a change in the definition of the CPI and in the different treatments applied, such as discounts and offers. However, taking into account the objective of the CPI and in view of the results that will be obtained with the use of the databases, a global vision of the evolution of market prices will be offered.

- **Differences between product and the full range of varieties**

The use of databases allows information on all varieties of a product sold. For example, suppose that for olive oil, in an establishment, there may be around 250 varieties (250 different EANs). This differs from the usual CPI procedure which, by its conception, tracks the prices of a single variety in each establishment.

- **Volume of information**

With scanner data the volume of data is significantly higher than that obtained with traditional price collection. Therefore, in addition to the technical requirements for the treatment of such information, methodological developments must also be introduced, to determine the varieties that must be part of the sample each month.

These conceptual issues do not imply insurmountable pitfalls that prevent the use of scanner data in the CPI, but entail the need to develop new systems for managing large databases, methodological designs for the treatment of information and new methods for validating the results.

¹ EAN: European Article Number (13 digits of the article barcode)
The next section focuses on these new methods to adapt the information received from companies to calculate the CPI.

3. Process of calculating price indices from the information generated by the *scanner data* method

- **Correspondence of classifications**

  The great challenge of *scanner data* is to obtain price indices compatible with those of the CPI from the information contained in the companies’ databases, since these indexes must be integrated with the functional groupings in which the CPI is structured, based on the ECOICOP.

  Therefore, the first step necessary for the use of *scanner data* is to find the *correspondence between the classification of products used internally by the partner company and the classification of ECOICOP articles used in the IPC*. This assignment is essential to automatically associate the varieties with the items to which they refer only through the family of products of which it is a part.

- **Selection of articles**

  As in the traditional collection CPI, a sample of representative items must be selected within each ECOICOP plot, whose price evolution will represent the evolution of the plot.

  In the case of *scanner data*, this basket of items will be made based on the income of the previous year, discarding those items whose consumption is not representative. This basket of items will remain fixed for a whole year, and will be reviewed at the end of each year, including items that have begun to be consumed during the year and excluding those that have stopped.

  Each chain of establishments included in the calculation of the CPI will have its basket of representative *data scanner* items, depending on the items sold in that chain.

- **Selection of Varieties**

  Although one of the principles underlying *scanner data* is the maximum use of all the information available in the companies’ databases, the practice indicates that some of this information may not be useful for the purpose of the project, and may even cause a delay in the calculation process. These are varieties whose sales are not significant and/or are unstable over time, which makes them unrepresentative of the overall evolution of the prices of the item, requiring correction methods so that their entry or exit from the market does not condition the index evolution.

  Therefore, it is necessary to select the best-selling varieties of each item to be part of the sample. This selection can be done in two different ways, in a **static way**, that is, the selection is made once a year, with the information from the previous year (as for the selection of the articles) and remains stable throughout the year, or a **dynamic sample** can be established, where month by month, the sample of varieties that will enter the CPI calculation is selected.
The static method is based on the maintenance of a fixed variety basket. This has the consequence that for varieties that cease to be marketed throughout the year, imputations of their unit value must be made until the beginning of the following year, when the new variety selection will be made. Likewise, although the varieties were stable throughout the year, with the static method there may be cases in which varieties that have ceased to be representative in expenditure are maintained in the sample and, on the contrary, new ones with higher consumption are not incorporated. The great potential of *scanner data* would be wasted.

On the other hand, the dynamic method, in which the varieties selected for the calculation can vary between two consecutive months, automatically eliminates those products that are no longer representative and incorporates the new ones, provided that the magnitudes of the income are significant. And it combines the representativeness of the basket with the comparability over time, since it calculates the monthly variations with the varieties in both months.

As we have already pointed out above, each article is made up of numerous varieties whose identification is their respective EAN code. The process to select which products are part of the calculation and which must disappear each month is carried out independently in each of the items, for each of the establishment formats (hypermarket, large supermarket, small supermarket, etc.) that the chain has, and in each province.

- **CPI Scanner data calculation formula**

The procedure for calculating the CPI of *scanner data* is described below, from the selection of varieties of each product to its aggregate index. Until the aggregate indexes are obtained, this process is carried out for each of the varieties, establishment format and province.

1) **Selection of Common Varieties**

From the aggregated data of the first three full weeks of the reference month $m$, the common varieties are selected between month $m$ and the previous month $m-1$.

Note: Throughout this section, when mention is made of month $m$, it refers to those first three full weeks of said month.

2) **Calculation of the percentages of income of each variety within the article**

For each of the selected varieties as described in the previous section, the weight is calculated, as a percentage, that each variety has within the article to which it belongs, both for month $m$ and for $m-1$:

\[
Peso\ variedad_{i,f,p}^{m,t} = \frac{Ingreso\ variedad_{i,f,p}^{m,t}}{\sum_{i \in j} Ingreso\ variedad_{i,f,p}^{m,t}} * 100,
\]
where:

\[ \text{Ingreso variedad}_{i,f,p}^{m,t} \] are the income, in euros, of the \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

### 3) Selection of the varieties that will be used to calculate the price evolution of the scanner data item

To ensure that the varieties that will be part of the calculation are representative of consumption, among the common varieties in the two months, only those that have a significant percentage of sales will be used for the calculation. In addition, the selected varieties must cover a high percentage of the total sales of the item.

Thus, for the selection of varieties, the following algorithm will be used:

\[
\frac{\text{Peso variedad}_{i,f,p}^{m-1,t} + \text{Peso variedad}_{i,f,p}^{m,t}}{2} > \frac{1}{N_{f,p} \times \delta} \times 100,
\]

where:

\( \text{Peso variedad}_{i,f,p}^{m-1,t} \) is the weight of the \( i \)-th variety, common between \( m \) and \( m-1 \), of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m-1 \) of year \( t \),

\( \text{Peso variedad}_{i,f,p}^{m,t} \) is the weight of the \( i \)-th variety, common between \( m \) and \( m-1 \), of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \),

\( N_{f,p} \) is the number of varieties of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), common between months \( m \) and \( m-1 \) of year \( t \), and

\( \delta = 1.25 \), which makes the threshold determined by the formula select common varieties so that a percentage of sales greater than 80% is covered.

In addition, the varieties selected in this way will be added to those that were part of the the previous month’s sample but have not been selected in the current one, either because in month \( m \) they have or have been been sold, but not enough to exceed the threshold.

If after three months, the variety still does not exceed the selection expense threshold or disappears, the variety will be permanently removed from the basket. That is, once a variety is selected one month because it has had a significant expense, it will remain in the sample for three consecutive months even if it has no consumption. In this way, it is sought to avoid discarding varieties whose consumption has decreased or disappeared in a timely manner.

### 4) Unit value calculation (VU) of each variety

For each variety included in the sample, the unit value is calculated as the ratio between the income and the number of units sold:
\[
VU\text{ variedad}_{i,f,p}^{m,t} = \frac{\text{Ingresos variedad}_{i,f,p}^{m,t}}{\text{Unidades vendidas variedad}_{i,f,p}^{m,t}},
\]

where:

- \( VU\text{ variedad}_{i,f,p}^{m,t} \) is the unit value, in euros, of the \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

- \( \text{Ingresos variedad}_{i,f,p}^{m,t} \) are the income, in euros, of the \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

- \( \text{Unidades vendidas variedad}_{i,f,p}^{m,t} \) are the units sold of the variety \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

It is important to note that this unit value does not have to coincide with the purchase price of the product, since it does not correspond to a single transaction, but to all those that have occurred over the three weeks in which said product has been able to have different prices.

5) Calculation of the monthly variation of the article

From the unit values of the varieties that make up the basket of month \( m \), the monthly variation of the article will be calculated, as follows:

a. The **geometric mean of the unit values of month** \( m \) of year \( t \), of the varieties selected in month \( m \) of year \( t \) is calculated:

\[
\overline{VU}_{j,f,p}^{m,t} = \sqrt[n_{j,f,p}^{m,t}]{\prod_{i=1}^{n_{j,f,p}^{m,t}} VU\text{ variedad}_{i,f,p}^{m,t}},
\]

, where:

- \( \overline{VU}_{j,f,p}^{m,t} \) is the average of the unit values of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

- \( VU\text{ variedad}_{i,f,p}^{m,t} \) is the unit value of the \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

- \( n_{j,f,p}^{m,t} \) is the number of varieties selected of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

b. The **geometric mean of the unit values of month** \( m \) of year \( t \), of the varieties selected in month \( m \) of year \( t \) is calculated:

\[
\overline{VU}_{j,f,p}^{(m-1),t} = \sqrt[n_{j,f,p}^{m,t}]{\prod_{i=1}^{n_{j,f,p}^{m,t}} VU\text{ variedad}_{i,f,p}^{(m-1),t}},
\]

, where:
\( \overline{VU}_{j,f,p}^{(m-1),t} \) is the average of the unit values of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

\( VU_{\text{variedad}}_{j,f,p}^{(m-1),t} \) is the unit value of the \( i \)-th variety of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

\( n_{j,f,p}^{m,t} \) is the number of varieties selected of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

As can be seen in the formulation, the varieties that intervene in the geometric means of the two months are the same, since the variation of the article will thus only reflect variation of prices and not variation due to the different composition in the basket of varieties.

c. The **monthly variation of the article** is calculated, as well as the variation of the average unit values:

\[
\Delta_{j,f,p}^{m,t} = \left( \frac{\overline{VU}_{j,f,p}^{m,t}}{\overline{VU}_{j,f,p}^{(m-1),t}} - 1 \right) \times 100
\]

, where:

\( \Delta_{j,f,p}^{m,t} \) is the variation of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

\( \overline{VU}_{j,f,p}^{m,t} \) is the unit value of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

\( \overline{VU}_{j,f,p}^{m-1,t} \) is the unit value of scanner data item \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \).

It should be noted that the average unit values of month \( m-1 \), used for calculations of the monthly variation of month \( m \), will be different from that used to calculate the monthly variation of month \( m-1 \) (since in the latter, the selected varieties were the common ones between \( m-1 \) and \( m-2 \)).

6) **Calculation of Indices referring to December of the previous year**

In the field of CPI, an elementary aggregate is the consumption component with the lowest level of aggregation for which indices are obtained, and in whose calculation weights are not involved. The indices of these aggregates are called elementary indices.

The CPI calculates a basic index for each article of the shopping basket in each of the provinces, and therefore, the basic aggregate is the article-province.

Similarly, in the case of **scanner data**, an elementary aggregate represents a **scanner data** item, in a particular establishment format and in a specific province, that is, **article-format-province**. Therefore, the calculation formula of an elementary index is as follows:
\[
dic(t-1)_{j,f,p}^{m,t} = dic(t-1)_{j,f,p}^{(m-1),t} \times \left( 1 + \frac{\Delta_{j,f,p}^{m,t}}{100} \right)
\]

, where:

\(dic(t-1)_{j,f,p}^{m,t}\) is the index of scanner data item \(j\), of the establishment format \(f\) and province \(p\), in month \(m\) of year \(t\), refering to December of year \(t-1\),

\(dic(t-1)_{j,f,p}^{(m-1),t}\) is the index of scanner data item \(j\), of the establishment format \(f\) and province \(p\), in month \(m-1\) of year \(t\), refering to December of year \(t-1\), and

\(\Delta_{j,f,p}^{m,t}\) is the monthly variation of scanner data item \(j\), of the establishment format \(f\) and province \(p\), in month \(m\) of year \(t\).

Note: unlike the calculation of the traditional CPI, in which the elementary index can be expressed as a quotient of the average prices of the current month and that of December \(t\), the CPI with scanner data items is calculated from the monthly variation rates because, as already stated, the content of the basket of varieties of each article is homogeneous only between each pair of consecutive months.

7) **Calculation of aggregate indices referring to December of the previous year**

Calculation of weightings
The weights of each scanner data item represent the relationship between the expense incurred in that item and the total expense incurred on all items sold by the chain that are part of the index.

Therefore, the weights in force during year \(t\) will be calculated based on the annual sales of each item, referring to the previous year, taking into account all the weeks of the year (not the three monthly weeks used for calculations of the monthly variations of the IPC scanner data ), as follows:

\[
W_j^t = \frac{\text{ingresos del artículo } j \text{ en el año } (t - 1)}{\text{ingresos totales en el año } (t - 1)}
\]

These weightings are different for each establishment format in each of the geographical aggregations (provinces, Autonomous Communities, and the total national), and from them, the weightings of the different functional aggregations are obtained. Thus, the weighting of functional aggregate \(A\) is obtained as the sum of the weightings of the scanner data articles that comprise said aggregation:

\[
W_A^t = \sum_{j \in A} W_j^t
\]

Functional aggregations within a province
The index, referring to December of the previous year, of a functional aggregation \(S\) (5 digits ECOICOP) in a province \(p\), is obtained as an aggregation of the elementary indexes of the articles belonging to said aggregation, in each of the formats, with the Weights in force in year \(t\):
where:

\[ dic(t-1)I_{S,p}^{m,t} = \sum_{j \in S} \sum_{f=1}^{3} dic(t-1)I_{j,f,p}^{m,t} \times W_{j,f,p}^t \]

\[ I_{j,f,p}^{m,t} \] is the index referring to December of (t-1) of article \( j \), of the establishment format \( f \) and province \( p \), in month \( m \) of year \( t \), referring to December of the year (t-1),

\[ W_{j,f,p}^t \] is the weighting of item \( j \), of the establishment format \( f \) and province \( p \), in force during year \( t \), and

\( S \) are the 5-digit ECOICOP subclasses.

With these aggregations we can obtain the indexes of any scanner data item, adding the establishment formats of the chain.

Likewise, we can obtain the indexes of an ECOICOP subclass in a certain format, adding the indexes of the articles of that format, which belong to that subclass:

\[ dic(t-1)I_{f,S,p}^{m,t} = \sum_{j \in S} I_{j,f,p}^{m,t} \times W_{j,f,p}^t \]

Geographical aggregations of a functional aggregation

In the same way as the previous case, the calculation of the index of a geographical aggregation \( R \) greater than the province, for a specific functional grouping \( A \) is calculated as follows:

\[ dic(t-1)I_{A,R}^{m,t} = \sum_{p \in R} dic(t-1)I_{A,p}^{m,t} \times W_{A,p}^t \]

Where:

\[ I_{A,p}^{m,t} \] is the index referring to December of (t-1) of functional group \( A \) in province \( p \), and

\[ W_{A,p}^t \] is the weighting of functional aggregation \( A \) in province \( p \), in force during year \( t \).

Integration of the traditional collection CPI with the scanner data CPI

The integration of the scanner data IPC with the traditional collection IPC will be done at the CPI subclass/province level, since it is the minimum common aggregation between the traditional collection basket and the scanner data basket. For the aggregation of these sets of indices it is necessary to have the relative weights of both.

Calculation of weightings

From the weights of the CPI in force every year, a new structure must be calculated in which each supermarket chain will have its own weight in each of the provinces.

This weight will be the same in all subclasses in which the scanner data will be incorporated and in their respective elementary aggregates. The weighting part of the subclass/province corresponding to the chain will be divided between the scanner data elementary aggregates and the rest of the weighting between the elementary aggregates of traditional collection, proportionally.
Therefore, the weights of the traditional collection CPI + scanner data are calculated as follows:

- For each article \( j \) of the scanner data basket, within subclass \( S \):

\[
s_{SC} \tilde{W}_{j,f,p}^t = r_{RT} W_{S,p}^t * \text{coef}_p^C * \frac{s_{SC} W_{j,f,p}^t}{\sum_{j \in S} s_{SC} W_{j,f,p}^t},
\]

where

\( s_{SC} \tilde{W}_{j,f,p}^t \) is the weighting of the scanner data item \( j \), of the establishment format \( f \) and province \( p \), in force during year \( t \), within the traditional collection CPI + scanner data,

\( r_{RT} W_{S,p}^t \) is the weighting of subclass \( S \) of the traditional collection CPI, in province \( p \), in force during year \( t \),

\( \text{coef}_p^C \) is the weight, for one, of the chain \( C \) compared with traditional collection, in the province \( p \), and

\( s_{SC} W_{j,f,p}^t \) is the weighting of data scanner item \( j \), of the establishment format \( f \) and province \( p \), in force during year \( t \).

- For each article \( j \) of the scanner data basket, within subclass \( S \):

\[
r_{RT} \tilde{W}_{k,p}^t = r_{RT} W_{S,p}^t * \left( 1 - \text{coef}_p^C \right) * \frac{r_{RC} W_{k,p}^t}{\sum_{k \in S} r_{RC} W_{k,p}^t} = r_{RC} W_{k,p}^t * \left( 1 - \text{coef}_p^C \right),
\]

where

\( r_{RT} \tilde{W}_{k,p}^t \) is the weighting of the traditional collection item \( k \), in province \( p \), in force during year \( t \), within the traditional collection CPI + scanner data,

\( r_{RT} W_{S,p}^t \) is the weighting of subclass \( S \) of the traditional collection CPI, in province \( p \), in force during year \( t \),

\( \text{coef}_p^C \) is the weight, for one, of the chain \( C \) compared with traditional collection, in the province \( p \), and

\( r_{RC} W_{k,p}^t \) is the weighting of the traditional collection item CPI, in province \( p \), in force during year \( t \).

In this way a weighting structure is obtained in which the weighting of subclass \( S \) in province \( p \) will be the sum of the traditional collection weighting and the scanner data weighting:

\[
r_{RC+SC} \tilde{W}_{i,p}^t = \sum_{j \in S} \sum_{f} s_{SC} \tilde{W}_{j,f,p}^t + \sum_{k \in S} r_{RT} \tilde{W}_{k,p}^t
\]
Calculation of aggregate indices referring to December of (t-1)

With this new weighting structure and with the elementary indexes, both of data scanner and traditional collection, functional aggregate indices referring to December of (t-1) are obtained.

The index referring to December of the previous year, of a functional aggregation A in a province p in month m of year t will be calculated as follows:

\[ dic(t-1)J^{m,t}_{A,p} = \sum_{j \in A} \sum_{f} dic(t-1)J^{m,t}_{j,f,p} * SC\bar{W}^{t}_{f,f,p} + \sum_{k \in A} dic(t-1)J^{m,t}_{k,p} * RT\bar{W}^{t}_{k,p} , \]

where

- \( dic(t-1)J^{m,t}_{j,f,p} \) is the index of scanner data item j, of the establishment format f and province p, in month m of year t, referring to December of year t-1,

- \( SC\bar{W}^{t}_{f,f,p} \) is the weighting of the scanner data item j, of the establishment format f and province p, in force during year t, within the traditional collection CPI + scanner data,

- \( dic(t-1)J^{m,t}_{k,p} \) is the index of the traditional collection article k, in province p, of month m of year t, referred to December of year (t-1),

- \( RT\bar{W}^{t}_{k,p} \) is the weighting of the traditional collection item k, in province p, in force during year t, within the traditional collection CPI + scanner data,

Calculation of indices in base 2016

The indices referring to December of (t-1) are linked to the corresponding indices in base 2016. Thus, the base 2016 index of a functional aggregation A in a province p in month m of year t will be:

\[ 16^{1m,t}_{A,p} = 16^{1dic(t-1)}J_{A,p} * \frac{dic(t-1)J^{m,t}_{A,p}}{100} \]

For \( t = 2019 \), the \( 16^{1m,t}_{A,p} \) are those calculated only with traditional collection. For \( t = 2020 \), the \( 16^{1dic19}_{A,p} \) will therefore be calculated with traditional collection and \( dic(t-1)J^{m,t}_{A,p} \) those calculated as a union of the traditional collection CPI and the scanner data IPC.

The indices for the different geographical aggregations will be calculated analogously.