Chapter 6 - Calculation of Indices: Aggregation

Chapter 5 aimed to set out how individual data are combined to form elementary indices, the first true level of computation and analysis of IPI indices. These indices provide an estimate, for each of the families of industrial products monitored, of the quantities produced on a monthly basis by enterprises located in France. They are then aggregated based on different levels of classification to facilitate the analysis and use of the indicators.

1- General Principles of the Aggregation of Indices with Base Year 2015

1.1- From Constant Weights to Chained Indices with Annually Updated Weights

The transition of industrial production indices to the 2015 base year, first published in March 2018, was accompanied by the introduction of annual chain linking, using a set of annual weights for aggregating the indices at higher levels. In previous bases, fixed weights were used to aggregate IPI series over the entire 5-year calculation period (2005 weights for the 2005 base, 2010 weights for the 2010 base, etc.).

The new methodology improves the long-term robustness of the index by:

- taking better account of structural changes in the economy (in practical terms, if, at the most detailed level – i.e. elementary series – changes depend only on the specific data collected for the IPI, "regular calibration" on structural statistics limits any long-term drifts);
- taking into account the distortion of relative prices (similar to what is done in the national accounts in calculating volumes based on the previous year's chained prices, such as GDP, which helps to improve the consistency between the IPI and output calculated by national accounts on the basis of structural statistics);
- lastly, it significantly reduces revision problems during base changes[29].

While this change appears to provide increased quality for the IPI, it is important not to overestimate the impact of the change in weights each year: in particular, since changes in the productive structure and in relative prices are relatively slow, aggregating indices from weights associated with one year or those of the previous year would, in most cases, only marginally affect the changes in the indices.

Conversely, using chained indices has some disadvantages related in particular to the increased complexity (see below).

1.2- Annual Overlap Chaining

There is no single method for chaining indices or, more generally, quantities based on the quantities available at the lower levels. The method used for the IPI is known as the "annual overlap" method. The construction of the index for a given month depends directly on the average annual index for the previous year and on a trend calculated based on carefully selected sub-indices and weights. The definition and specific construction of the weights is the subject of the following section.

Let:
- m be the month in which the index is to be calculated and A be the year;
- $Aref$ be the reference year (=2015 currently);

29In systems using constant weights, the change in reference year for the calculation of weights implied updating the entire series based on the new weights, which could lead to significant and sometimes incomprehensible revisions.

- $J^{m,A}_{Agg}$ = Index of aggregated series Agg for reference month m of year A;
- J^{A}_{Agg} = Annual average of the indices of the aggregate series for year A;
- $J^{m,A}_{i}$ = Index of subseries i (lower level), such that $i \in Agg$, for reference month m of year A and J^{A}_{i} be the annual average of the index for series i for year A;
- p^{A}_{i} be the weighting associated with series i ($i \in Agg$), used for the calculation of year A (typically a variable depending on the value added of the previous year, see next chapter).

The general formula for calculating the indices is then written[30] : $J^{m,A}_{Agg} = J^{A-1}_{Agg} * \frac{\sum_i p^{A}_{i} J^{m,A}_{i}}{\sum_i p^{A}_{i} J^{A-1}_{i}}$

We can then easily understand the construction method: we start from the level of the index of the previous year and construct the monthly aggregated indices of year A with reference to this annual index. To achieve this, a change index is applied based on the weighted averages of the sub-indices (monthly in the numerator and annual in the denominator). The index is:

- chained[31] since the index is deduced from the value of the previous period and so on;
- with annual overlap, since the chaining is based on the annual average of the previous year, thereby ensuring the quality of the annual change and providing a basis for moving directly to a chained annual index by summing the monthly indices.

Indeed, if the indices of the different months of year A are summed up, we easily obtain $\frac{J^{A}_{Agg}}{J^{A-1}_{Agg}} = \frac{\sum_i p^{A}_{i} J^{A}_{i}}{\sum_i p^{A}_{i} J^{A-1}_{i}}$. In other words, the annual change of the aggregate index is obtained as the ratio of the averages of the sub-indices weighted by the same system (the weights $p^{A}_{i}, i \in A$; see Chapter 7 for a presentation of the calculation of these weights): the index obtained is a (chained) Laspeyres index.

30The formula should be adjusted slightly to calculate the indices over the period prior to the base year (calculations are made on the basis of the indices of that base year, before and after) and for the base year serving as a reference.
31The IPI is therefore ultimately a double-chained index:
- it will be derived from a calculation of elementary indices by chaining the series month to month;
- the aggregated indices are also calculated by chaining from the previous year's indices and the lower level indices.

Comments:

- The aggregation formula for indices with constant weights was written: $J_{Agg}^{m,A} = \frac{\sum_i p_i^{Aref} J_i^{m,A}}{\sum_i p_i^{Aref}}$

- The implementation of the new methodology is accompanied by backcasting of the weights over a long period (see below);

- Using chaining methods for the IPI has some disadvantages related in part to the greater complexity of the calculations involved:
    - the exact calculation of the contributions of subseries to changes in an aggregate is more complex;
    - the change in the first month of the year relative to the last month of the previous year is affected by the change in weight (see Section 3).

2- The Principle Governing the Calculation of Weights

2.1- Type of Weights Used

The European Regulation on short-term statistics[32] specifies, for each short-term indicator, the preferred weighting variable to be used for the aggregation of elementary indices. In the case of the IPI, the weighting variable is value added (VA). It is important to note that the choice of VA is based on practical and theoretical considerations as well as on conventions (see box). As noted in Chapter 1, at the elementary level, the changes measured by the IPI are not exactly similar to changes in value added.

Eurostat recommends using the VA data drawn from structural business surveys. However, in France, structural business statistics do not allow for a direct calculation of VA per branch. The calculation of the weights is first based on the VA data by branch contained in the national accounts (aggregate classification, level A129). However, this level is not detailed enough for the IPI. A breakdown to the most detailed level is then carried out based on annual structural data on turnover by branch (see diagram).

Box 1: Why Use Value Added Weights?

The use of valued added as a weight unit for aggregating indices is consistent with a Eurostat recommendation. This recommendation implies a particular measure of changes in activity.
The first option is to construct Laspeyres indices of quantities produced using price weights. These are the simplest indices and appear to be consistent with the standard representation in terms of the value of "baskets" of quantities produced. However, they have the disadvantage of valuing the production of intermediate goods several times over.
Industrial production as a whole consists of products, some of which – i.e. consumer goods and capital goods – have reached the final stage of development and are therefore no longer processed until they reach their final destination, while others – i.e. intermediate goods – must be processed several times before reaching the final stage of development. Gross production indices are relevant for monitoring the production of consumer goods or capital goods: the system of weighting using base year unit prices is clear and fairly appropriate for valuing a basket of products, all of which have reached the final stage of development. On the other hand, gross production indices are not suitable for monitoring the production of intermediate goods or for monitoring the production of manufacturing as a whole since they overweigh the production of goods downstream of the production process. For example, the use of a gross production index values the production of the "automotive" series at a price that includes the value of the steel sheets, glass, tyres, etc. used to manufacture it, despite the fact that the production of these goods has already been included in the "steel sheets", "flat glass" and "tyre" series. In doing so, the value of intermediate goods is counted several times. The weighting system does not give each product family a weight proportional to the economic importance of the production of these products: it is biased insofar as it systematically assigns a greater weight to products that are located downstream of the production chains.

32*See Methodology of short term business statistics - Interpretation and Guidelines, Eurostat, 2006*

The concept of VA can be used to overcome this difficulty. Value added is the balancing item of the production account. It is calculated as the value of production less intermediate consumption.
This index is called the net production index to underline the choice of a weighting system that distinguishes it from the simpler gross production index. In practice, however, the term “net” is generally omitted.

2.2- Data Used and Calculation of the Distribution of VA

2.2.1-National Accounts Data

To strike a balance between data availability and data robustness, under the current system the calculation of weights is based on the semi-definitive version of the national accounts (available in May N+2 for year N). However, the weights used for backcasting (chained indices with annually updated weights) over a long period are based on the final version of the accounts.

The variable used is **value added at base prices of industrial branches across all institutional sectors**[33]. VA at base prices is defined as value added at producer prices minus any tax payable, and plus any subsidy receivable:

VA at base prices = gross VA[34] + Subsidies– taxes on products

2.2.2-ESANE Data

Individual Data or Composite Aggregates

The breakdown at the subclass level of national accounts data is carried out based on the turnover data[35] provided by the structural business statistics obtained from ESANE. ESANE data are available as individual data files, as well as in aggregate form at the subclass level. Known as composite aggregates, these aggregates are obtained after a complex adjustment process. In both cases, there are “sector” data (data collected at the “enterprise” level) and “branch” data (breakdown by branch based on turnover).

The calculation of the weights is now based on aggregated files. This seems to be more appropriate given that ESANE macro controls are applied to the aggregates, thereby ensuring greater reliability. The implementation process is also quicker.

We therefore have:

$$Weight_{subclass\ i} = Value\,added^{national\ accounts}_{A129\ level \ni i} \times \frac{Turnover^{Esane}_{sous\text{-}classe\ i}}{Turnover^{Esane}_{A129\ level \ni i}}$$

33In manufacturing, there are only a small number of branches where the value added of non-financial corporations and sole proprietorships differs from the value added of institutional sectors as a whole. These are primarily branches E36Z (Water collection, treatment and supply) and E37Z (Sewerage), where communal bodies and public companies are common. For the sake of simplicity and consistency with what is published, and as was the case for the previous bases, a decision was made to retain the added value on all institutional sectors.

34VA = Sales of goods - Purchases of goods - Changes in stocks of goods
+ Output of goods sold
+ Output of sold services
+ Output sold as inventory
+ Capitalised output
+ Other production
- Purchases of raw materials
+ Changes in stocks of raw materials
+ Other purchases and external charges
+ Other costs.

35VA data by branch being unavailable.

2.2.3-EAP Data

Annual Production Survey (EAP) data are used to break down VA at the elementary series level (see Figure 1).

The process of collecting data on industrial products is conducted at a very detailed level of classification. As such, this tool is particularly suitable for calculating weights at the most detailed level of the elementary series, which are moreover derived from the EAP product classification.

The variable used corresponds to “invoicing” in the EAP. Since the definition of industry used in the IPI is based on the UN definition, model M1 (see Chapter 2) was excluded for all products for which a breakdown by economic model is available in the survey. For other products or activities, total invoicing was used.

The EAP is used for all branches of industry except for food and agriculture, where the breakdown is based on the annual branch surveys of the SSP (Statistical Service of the Ministry of Agriculture), or in cases where the boundaries of the elementary series coincide with the boundaries of the subclass (e.g. aeronautics).

We therefore have:

$$Weight_{elementary\ series\ i} = Weight_{sous\text{-}classe \ni i} \times \frac{Invoicing^{EAP}_{elementary\ series\ i}}{Invoicing^{EAP}_{subclass \ni i}}$$

Comment 1:

At the level of the elementary series, until 2016 the IPI weights did not use the total VA of a branch but the proportion of the VA covered by the IPI series (for a definition of the coverage rate, see Chapter 3). The choice to be made depends on the intended message or meaning to be given to the IPI dissemination.

- Case 1: The VA of the Branch is Adjusted by the Coverage Rate
For example 2, consider two branches A and B with the same VA (100). In Branch A, the IPI covers 70% while in Branch B the IPI covers 50%. If the index of A increases by 1 point while the index of B decreases by 1 point, the change in the aggregate (A+B) will be positive. Implicitly, the message to users is that branch A is more represented.

- Case 2: The VA of the Branch is not Adjusted by the Coverage Rate
In this case, the message is different, with the published IPI representing changes in the branch’s output regardless of coverage. In other words, it is assumed that changes in the IPI over an elementary series reflect changes in the entire associated branch, despite the coverage of manufactured products being below 100%. Using the example above, the change in the aggregation (A + B) would be zero.

INSEE has opted for the second solution, which seems more relevant and more in line with the IPI’s publication objectives.

Comment 2:

For the “construction” sector, Eurostat has developed a specific classification, the Classification of types of construction (abbreviated as CC), which is not directly related to NACE-4. A correspondence table[36] between the NAF Rev.2 subclass codes and the aggregates of this classification, namely “buildings” and “civil engineering works”, was therefore developed.

Within the scope of the forthcoming European Regulation on short-term statistics, the production index for construction will be based on NACE and will distinguish between NACE Divisions 41, 42 and 43.

36It should be noted that a direct relationship existed between NAF Rev. 1 and CC.

2.2.4-Importance of Adjusting the VA Weights and Recalibration

Since the weights are applied to indices and not to changes (see the formula presented in 1.2), the weights in VA must be adjusted by the changes in the indices since the base year. It is important not to double count changes in volume between the reference year and the base year, i.e. once in the weights and a second time in the indices.

Consider the example of two indices A and B aggregated to level C, the two branches being equally distributed (each representing 50% of series C series in value-added terms). By definition, the two indices A and B equal 100 on average in 2015. We now position ourselves in 2017 and assume that branch A has doubled in volume in the meantime while branch B has remained at the same level. The fact that branch A has doubled in size should appear in the VA data from the national accounts and used to calculate the weights. Excluding price effects, the VA of branch A now represents 67% of the total (branch C) while branch B accounts for 33%. At the same time, index A doubled, increasing to 200, while index B remained constant at 100. We can see that if we aggregate indices A and B in 2017 based on the weights in 2017 VA, the change observed in branch A will be counted twice: once because the index has doubled (normal) and a second time through the update of its weighting. It is therefore necessary to adjust the 2017 weight based on the change in value added between 2015 and 2017.

Using the notations of Section 1.2 and using va_i^{A-1} to denote the weight in VA terms of branch i in the next highest aggregate (*"Agg"*), the adjusted weights are then written as follows (regardless of classification level):

$$p_i^A = va_i^{A-1} * \frac{J_i^{Aref}}{J_i^{A-1}} = va_i^{A-1} * \frac{100}{J_i^{A-1}}$$

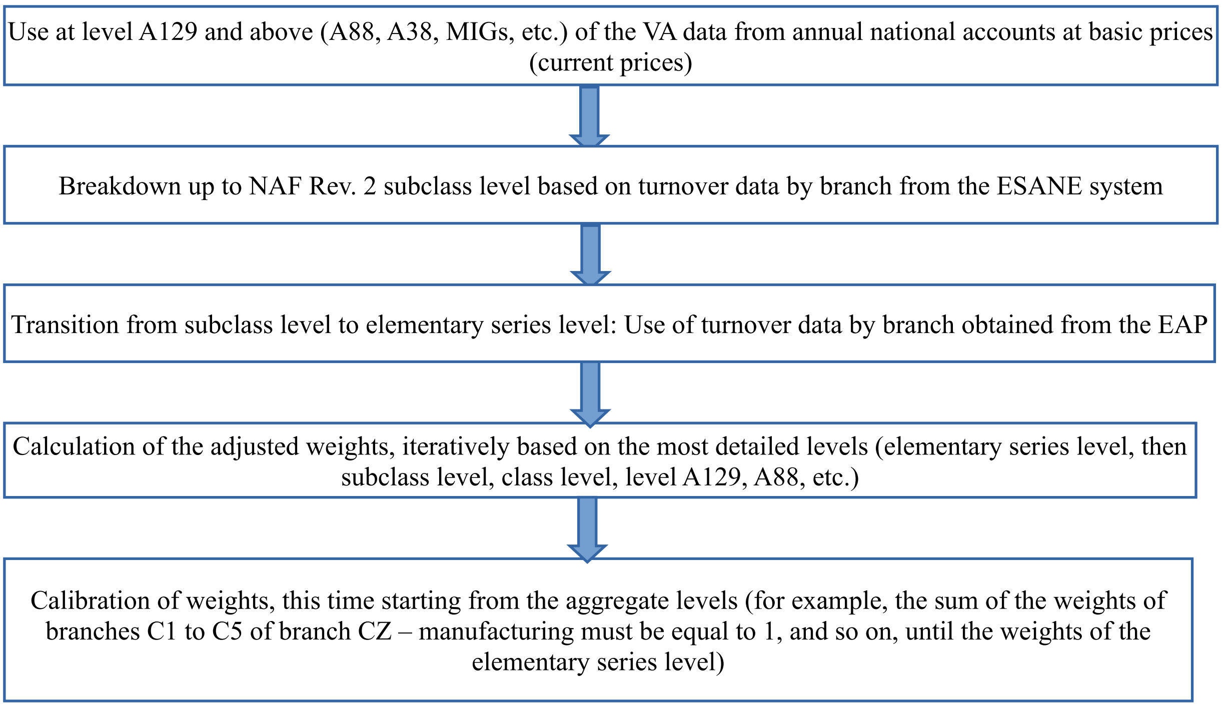
Since the annual average aggregate index of the previous year must be known for the calculation, it is necessary to progress iteratively from the most detailed levels (as part of a bottom-up approach):

- we start from the known elementary indices (see previous chapter) and calculate the annual average indices at the elementary level, allowing for the weights at this level to be calculated;
- using the elementary indices and weights, the indices can be calculated at the higher level;
- again, the weights can be calculated using known VA data and the average annual indices;
- and so on, up to the most aggregated levels.

Finally, to adhere to the additivity properties, we proceed in the opposite direction to adjust the weights (top-down approach) so that the sum of the weights of the sub-branches of an aggregate is equal to 1. All the operations involved in calculating the weights are summarised in Figure 1.

Figure 1: Main Stages in the Calculation of Weights

2.2.5-Backcasting of Weights over the Long Term



Given the many sources required to calculate the weights, it was possible to calculate them in a standard way from 2010 onwards. Previously, changes in classifications or sources (e. g. the transition from the EAB to the EAP) did not allow for a systematic calculation in this way.

In order to be able to chain the indices over a long period (since 1990), it was necessary to backcast the information available: national accounts by branch available over a long period (level A129 since 1999 and level A88 before that), and at the more detailed levels, use of the weights of the old IPI bases and interpolation between two reference years. Before 2000, the "*infra* NAF A88" structure was thus constant.

2.2.6-Updating Weights and Revisions

Given the delays in the availability of data used to construct the weights (for example, the "semi-final" annual accounts for 2017 are published in May 2019), the recent indices (year N) rely, at least initially, on weights calculated on previous data, until the N-1 data (reference year used to calculate the weights) are published. These operations can of course lead to revisions on recent years.

3- Going Further: Interpretation of Chaining Formulas for Monthly and Annual Changes

By showing the value added data in the chaining formula, after simplification we obtain:

$$J_{Agg}^{m,A} = J_{Agg}^{A-1} \times \sum_i \left(\frac{VA_i^{A-1}}{\sum_k VA_k^{A-1}} \right) \frac{J_i^{m,A}}{J_i^{A-1}}$$

with VA_i^{A-1} as the value added of branch i for year A-1.

The change in the aggregate index between months (m) and (m+1) of the same year (m cannot be December) is then written:

$$\frac{J_{Agg}^{m+1,A}}{J_{Agg}^{m,A}} = \sum_i \frac{VA_j^{A-1}}{\sum_k VA_k^{A-1}} \times \frac{J_i^{m+1,A}}{J_i^{m,A}}$$

Therefore, the change simply corresponds to the average of the elementary changes, weighted by the share of sub-branches in the aggregate. Similarly, the annual change between (A-1) and (A) is written

$$\frac{J_{Agg}^{A}}{J_{Agg}^{A-1}} = \sum_i \frac{VA_j^{A-1}}{\sum_k VA_k^{A-1}} \times \frac{J_i^{A}}{J_i^{A-1}}.$$

In other words, to calculate the annual change, we start from the value-added structure of the previous year and apply the changes in the indices as an annual average.

The relationship between January of year A and December of year A-1 is more complex since it implicitly incorporates the shock associated with the weight change (change from weights based on VA A-2 to weights based on VA A-1 with the year change).