

Chapitre 6 - Le calcul des indices : l'agrégation

Le chapitre 5 visait à présenter comment les données individuelles étaient combinées pour former les indices élémentaires, premier vrai niveau de construction et d'analyse des indices IPI. Ces indices donnent une estimation, pour chacune des familles de produits industriels suivies, des quantités produites mensuellement par les entreprises implantées sur le territoire français. Ils sont ensuite agrégés selon différents niveaux de nomenclature pour faciliter l'analyse et l'utilisation de ces indicateurs.

1- Principes généraux de l'agrégation des indices en base 2015

1.1- Des pondérations constantes aux indices chaînés à pondérations annuelles

Le passage des indices de la production industrielle à la base 2015 dont la première publication a eu lieu en mars 2018 s'est accompagné de la mise en place d'un chaînage annuel des indices à l'aide d'un jeu de pondérations annuelles pour l'agrégation des indices au niveaux supérieurs. Dans les bases précédentes, on utilisait pour agréger les séries IPI des pondérations fixes sur l'ensemble de la période de calcul (pondérations 2005 pour la base 2005, 2010 pour la base 2010...).

La nouvelle méthodologie permet d'améliorer la robustesse de l'indice sur longue période à travers :

- une meilleure prise en compte des évolutions structurelles de l'économie (concrètement, si au niveau le plus fin – séries élémentaires, les évolutions ne dépendent que des données en propre collectées pour l'IPI, le « calage régulier » sur les statistiques structurelles limite les dérives éventuelles sur le long terme) ;
- la prise en compte de la déformation des prix relatifs (à l'image de ce qui est fait par les comptes nationaux pour le calcul des volumes au prix de l'année précédente chaînés, comme le PIB par exemple, ce qui renforce la cohérence entre l'IPI et la production calculée par les comptes nationaux à partir des statistiques structurelles) ;
- enfin, elle réduit nettement les problèmes de révision lors des changements de base²⁶.

Si ce changement apporte donc *a priori* un vrai gain de qualité pour l'IPI, il ne faut en revanche pas surinterpréter l'impact du changement de pondérations chaque année : en particulier, les changements de structure productive et l'évolution des prix relatifs étant relativement lents, agréger les indices à partir de pondérations associées à une année ou à celles de l'année précédente ne modifieraient dans la plupart des cas les évolutions des indices que de façon marginale.

À l'inverse, l'utilisation d'indices chaînés revêt quelques inconvénients associés notamment à une plus grande complexité (*cf. infra*).

1.2- Chaînage en recouvrement annuel

Il n'existe pas de méthode unique pour chaîner des indices ou plus généralement des quantités à partir des quantités disponibles aux niveaux inférieurs. La méthode utilisée pour l'IPI est dite à recouvrement annuel (« *annual overlap* »). La construction de l'indice d'un mois donné dépend en effet directement de l'indice annuel

²⁶ Dans les systèmes à pondérations constantes, le changement d'année de référence pour le calcul des pondérations conduisait à réactualiser l'ensemble de la série à l'aune de ces nouvelles pondérations, ce qui pouvait occasionner des révisions importantes et parfois peu compréhensibles.

moyen de l'année précédent et d'une évolution calculée à partir des sous-indices et de pondérations bien choisies. La définition et la construction spécifique des pondérations fait l'objet de la section suivante.

Soient :

- m le mois où l'on souhaite calculer l'indice et A l'année ;
- A_{ref} l'année de référence (=2015 actuellement) ;
- $J_{Agg}^{m,A}$ = Indice de la série agrégée Agg pour le mois de référence m de l'année A ;
- J_{Agg}^A = Moyenne annuelle des indices de la série agrégée pour l'année A ;
- $J_i^{m,A}$ = Indice de la sous-série i (niveau inférieur), tel que $i \in Agg$, pour le mois de ; référence m de l'année A et J_i^A la moyenne annuelle de l'indice de la série i pour l'année A ;
- p_i^A la pondération associée à la série i ($i \in Agg$), utilisée pour le calcul de l'année A (typiquement une variable dépendant de la valeur ajoutée de l'année précédente, cf. chapitre suivant).

Alors la formule générale de calcul des indices s'écrit²⁷ :

$$J_{Agg}^{m,A} = J_{Agg}^{A-1} * \frac{\sum_i p_i^A J_i^{m,A}}{\sum_i p_i^A J_i^{A-1}}$$

On comprend alors aisément le mode de construction : on part du niveau de l'indice de l'année précédente et on construit les indices agrégés mensuels de l'année A en référence à cet indice annuel. Pour y parvenir, on applique un indice d'évolution construit à partir des moyennes pondérées des sous-indices (mensuels au numérateur et annuel au dénominateur). L'indice est :

- chaîné²⁸ car l'indice se déduit de la valeur de la période précédente et ainsi de suite ;
- à recouvrement annuel, car le chaînage s'appuie sur la moyenne annuelle de l'année précédente, ce qui permet de garantir la qualité de l'évolution annuelle et de passer directement à un indice annuel chaîné en sommant les indices mensuels.

En effet, si l'on somme les indices des différents mois de l'année A , on obtient facilement que

$$\frac{J_{Agg}^A}{J_{Agg}^{A-1}} = \frac{\sum_i p_i^A J_i^A}{\sum_i p_i^A J_i^{A-1}} . \text{ Autrement dit, l'évolution annuelle de l'indice agrégé s'obtient comme le ratio des}$$

moyennes des sous-indices pondérées par le même système (les pondérations $p_i^A, i \in A$, cf. chapitre 7 pour la présentation du calcul de ces pondérations) : c'est un indice de Laspeyres (chaîné).

²⁷ La formule doit être ajustée légèrement pour le calcul des indices sur la période qui s'étend avant l'année de base (les calculs sont effectués en partant des indices de cette année de base, avant et après) et pour l'année de base qui sert de référence.

²⁸ L'IPI est donc au final un indice doublement chaîné :

- il résultera d'un calcul d'indices élémentaires par chaînage au mois le mois des séries ;
- les indices agrégés sont également calculés par chaînage à partir des indices de l'année précédente et des indices de niveaux inférieurs.

Remarques :

- La formule d'agrégation dans le cas d'indices à pondérations constantes s'écrivait :

$$J_{Agg}^{m,A} = \frac{\sum_i p_i^{Aref} J_i^{m,A}}{\sum_i p_i^{Aref}}$$

- La mise en place de la nouvelle méthodologie s'est accompagnée d'une rétopolation des pondérations sur longue période (*cf. infra*) ;
- L'utilisation des méthodes de chaînage pour l'IPI revêt quelques inconvénients, liés en partie à la plus grande complexité des calculs :
 - le calcul exact des contributions de sous-séries à l'évolution d'un agrégat est plus complexe ;
 - l'évolution du premier mois de l'année par rapport au dernier mois de l'année précédente est affectée par le changement de pondération (*cf. section 3*).

2- Le principe de calcul des pondérations

2.1- Type de pondérations utilisées

Le règlement européen sur les statistiques de court terme²⁹ indique pour chaque indicateur de court terme la variable de pondérations à utiliser de façon privilégiée pour l'agrégation des indices élémentaires. Pour l'IPI, il s'agit de la valeur ajoutée (VA). Il est important de noter que ce choix relève à la fois de considérations pratiques, théoriques mais aussi de conventions (*cf. encadré*). En effet, comme évoqué dans le chapitre 1, au niveau élémentaire les évolutions mesurées par l'IPI ne s'apparentent pas exactement à des évolutions en la valeur ajoutée.

Eurostat préconise l'usage de la VA issue des enquêtes structurelles d'entreprises. Néanmoins, en France, les statistiques structurelles sur les entreprises ne permettent pas de calculer directement une VA par branche. Le calcul des pondérations s'appuie d'abord sur les données de VA par branche des comptes nationaux (nomenclature agrégée, niveau A129). Ce niveau n'est toutefois pas assez fin pour l'IPI, une ventilation au niveau le plus fin est alors ensuite effectuée à partir des données structurelles annuelles de chiffre d'affaires par branche (*cf. schéma*).

Encadré 1: Pourquoi utiliser des pondérations en VA?

L'utilisation de la VA comme unité de pondération pour agréger les indices répond à une préconisation d'Eurostat. Cette préconisation induit une mesure particulière de l'évolution de l'activité.

Ainsi, une première possibilité serait de construire des indices de Laspeyres des quantités produites à partir de pondérations en prix. Ces indices sont les plus simples et apparaissent conformes à la représentation habituelle en matière de valeur de « paniers » de quantités produites. Ils présentent cependant le défaut de valoriser plusieurs fois la production des biens intermédiaires.

La production industrielle dans son ensemble est constituée de produits dont les uns – les biens de consommation et les biens d'équipement – ont atteint le stade final d'élaboration et ne sont donc plus transformés avant d'atteindre leur destinataire final, et les autres – les biens intermédiaires – doivent être transformés plusieurs fois avant d'atteindre le stade final d'élaboration. Les indices de production brute sont pertinents pour suivre la production des biens de consommation ou celle des biens d'équipement : le système de pondérations par les prix unitaires de l'année de base est clair, et convient équitablement pour valoriser un panier de produits qui ont tous atteint le stade final d'élaboration. En revanche, les indices de production brute ne conviennent, ni pour suivre la

29 Cf. *Methodology of short term business statistics - Interpretation and guidelines*, Eurostat, 2006

production des biens intermédiaires, ni pour suivre la production de l'ensemble de l'industrie, car ces indices surpondèrent la production des biens situés en aval du processus de production. Par exemple, l'utilisation d'un indice de production brute valoriserait la production de la série « automobile » à un prix qui comprend la valeur des tôles d'acier, du verre, des pneumatiques, etc. qui ont servi à sa fabrication alors que la production de ces biens a déjà été prise en compte dans les séries « tôles d'acier », « verre plat » et « pneumatiques ». La valeur de biens intermédiaires serait alors comptée plusieurs fois. Le système de pondérations n'accorde pas à chaque famille de produits un poids proportionnel à l'importance économique de la production de ces produits : il est biaisé parce qu'il accorde systématiquement un poids plus élevé aux produits qui sont situés en aval des filières de fabrication.

Pour contourner cette difficulté, il est possible d'utiliser le concept de VA. La VA est un des soldes du compte de production. Elle est égale à la valeur de la production diminuée de la consommation intermédiaire.

Cet indice est appelé indice de production nette, de manière à rappeler le choix d'un système de pondérations qui permet de le distinguer de l'indice plus simple de production brute. En pratique toutefois, le qualificatif « nette » est le plus souvent omis.

2.2- Données utilisées et calcul de la répartition de la VA

2.2.1- Les données des comptes nationaux

Afin d'équilibrer entre disponibilité et robustesse des données, le calcul des pondérations s'appuie en régime courant sur la version semi-définitive des comptes nationaux (disponible en mai N+2 pour l'année N). Les pondérations utilisées pour la rétropolation (indices chaînés à pondérations annuelles) sur longue période s'appuient néanmoins sur les comptes définitifs.

La variable utilisée est la **valeur ajoutée au prix de base des branches industrielles sur l'ensemble des secteurs institutionnels**³⁰. La VA au prix de base se définit comme la VA au prix du producteur de laquelle sont déduits les impôts sur les produits et ajoutées les subventions sur les produits :

$$\text{VA au prix de base} = \text{VA brute}^{31} + \text{Subventions} - \text{impôts sur les produits}$$

Les données Esane

Données individuelles ou agrégats composites

La ventilation au niveau sous-classes des données des comptes nationaux est effectuée à partir des données en chiffre d'affaires³² des statistiques structurelles d'entreprise issues d'Esane. Les données d'Esane sont disponibles sous forme de fichiers de données individuelles, ainsi que sous forme **agrégée au niveau sous-classes**. Ces agrégats, appelés agrégats composites, sont obtenus à l'issue d'un processus complexe de redressement. Dans les deux cas, existent des données « secteurs » (données collectées au niveau « entreprise ») et de données « branches » (ventilation en branches du chiffre d'affaires).

Le calcul des pondérations s'appuie désormais sur les fichiers agrégés. Ce choix semble plus efficient puisque des macro-contrôles Esane s'appliquent aux agrégats, ce qui assure une fiabilité *a priori* plus grande. En outre, la mise en œuvre est plus rapide.

30 Dans l'industrie, il n'existe que quelques branches pour lesquelles la valeur ajoutée des SNFEI (Sociétés non financières et entreprises individuelles) diffère de celle de l'ensemble des secteurs institutionnels. Il s'agit essentiellement des branches E36Z (Captage, traitement et distribution d'eau) et E37Z (Collecte et traitement des eaux usées où l'on trouve des communes et des régions. Le choix a été fait, par simplicité et cohérence avec ce qui est publié, et comme cela était le cas pour les bases précédentes, de retenir la valeur ajoutée sur l'ensemble des secteurs institutionnels.

31 VA =
+ Ventes de marchandises - Achats de marchandises - Variations de stocks de marchandises
+ Production de biens vendue
+ Production vendue de services
+ Production stockée
+ Production immobilisée
+ Autres productions
- Achats de matières premières
+ Variations de stock de matières premières
+ Autres achats et charges externes
+ Autres charges.

32 Les données en VA par branche n'étant pas disponibles.

On a donc :

$$Poids_{\text{sous-classe } i} = VA_{\text{niveau A129 } \ni i}^{\text{comptes nationaux}} \times \frac{CA_{\text{sous-classe } i}^{\text{Esane branches}}}{CA_{\text{niveau A129 } \ni i}^{\text{Esane branches}}}$$

2.2.2- Les données de l'EAP

Les données de l'enquête annuelle de production (EAP) sont utilisées pour ventiler la VA au niveau des séries élémentaires (cf. figure 1).

La collecte des données sur les produits industriels se fait à un niveau très fin de nomenclature ; cet outil est donc tout particulièrement adapté pour calculer les pondérations au niveau fin des séries élémentaires, qui sont de plus dérivées de la classification des produits de l'EAP.

La variable utilisée correspond aux « facturations » de l'EAP. La définition de l'industrie retenue dans l'IPI étant celle de l'ONU, le modèle M1 (cf. chapitre 2) a été exclu pour tous les produits pour lesquels la ventilation par modèles économiques est disponible dans l'enquête. Pour les autres produits ou activités, les facturations totales ont été retenues.

L'EAP est utilisée pour toutes les branches de l'industrie sauf pour les industries agroalimentaires (IAA), pour lesquelles la ventilation repose sur les enquêtes annuelles de branche du SSP (Service statistique du ministère de l'agriculture), ou dans les cas où le contour de la série élémentaire coïncide avec celui de la sous-classe (cas par exemple de l'aéronautique).

On a donc :

$$Poids_{\text{série élémentaire } i} = Poids_{\text{sous-classe } \ni i} \times \frac{Facturations_{\text{série élémentaire } i}^{EAP}}{Facturations_{\text{sous-classe } \ni i}^{EAP}}$$

Remarque 1 :

Au niveau des séries élémentaires, jusqu'en 2016 les pondérations de l'IPI n'utilisaient pas la VA totale d'une branche mais la part de la VA couverte par les séries IPI (cf. chapitre 3 pour la définition du taux de couverture). Le choix à faire dépend du message ou du sens que l'on veut donner à la diffusion IPI.

- Cas 1 : la VA de la branche est corrigée par le taux de couverture

Si l'on considère par exemple 2 branches A et B avec la même VA (100). Dans la branche A, l'IPI couvre 70 % ; dans la branche B, l'IPI couvre 50 %. Si l'indice de A augmente de 1 point alors que l'indice de B diminue de 1 point, l'évolution de l'agrégat (A+B) sera positive. Implicitement le message destiné aux utilisateurs est que la branche A est davantage représentée.

- Cas 2 : la VA de la branche n'est pas corrigée par le taux de couverture

Dans ce cas, le message est différent, l'IPI publié représente l'évolution de la production de la branche indépendamment de la couverture ; autrement dit, on fait l'hypothèse que l'évolution de l'IPI sur une série élémentaire reflète l'évolution de l'ensemble de la branche associée, en dépit d'une couverture des produits fabriqués inférieure à 100 %. En reprenant l'exemple ci-dessus, l'évolution de l'agrégation (A + B) serait nulle.

L'Insee a opté pour la deuxième solution qui semble plus pertinente et plus conforme avec les objectifs de publication de l'IPI.

Remarque 2 :

S'agissant du secteur de la « construction », Eurostat a défini une nomenclature spécifique, la nomenclature CC (classification des types de construction), qui n'est pas directement reliée à la nomenclature NACE4. Une table de passage³³ entre les codes sous-classes en NAF rev.2 et les agrégats de cette classification, à savoir « bâtiment » et « travaux publics », a donc été construite.

Dans le cadre du prochain règlement européen sur les statistiques de court terme l'indice de la production dans la construction s'appuiera sur la nomenclature NACE et distinguera les niveaux 41, 42 et 43 de la NACE.

2.2.3- Nécessité de corriger les poids en VA et recalage

Les pondérations étant appliquées aux indices et non aux évolutions (*cf.* formule présentée en 1.2), il est nécessaire de corriger les poids en VA par l'évolution des indices depuis l'année de base. Il s'agit de ne pas compter 2 fois les évolutions en volume entre l'année de référence et l'année de base : une première fois dans les pondérations et une seconde fois dans les indices.

Prenons l'exemple de deux indices A et B que l'on agrège en un niveau C, les deux branches étant réparties équitablement (chacune représentant 50 % en part de VA de la série C). Par définition, les deux indices A et B valent 100 en moyenne 2015. On se place désormais en 2017 et supposons que la branche A ait doublé en volume entre-temps alors que la branche B soit restée au même niveau. Ce doublement d'importance doit apparaître dans les données de VA issues des comptes nationaux et utilisées pour le calcul des pondérations. Hors effets prix, la VA de la branche A représente désormais 67 % du total (branche C) et la branche B 33 %. Dans le même temps, l'indice A a doublé passant à 200 tandis que l'indice B est resté constant à 100. On voit donc que si l'on agrège les indices A et B en 2017 en s'appuyant sur les poids en VA 2017, on va compter deux fois l'évolution observée sur la branche A : une première fois car l'indice a doublé (normal) et une deuxième fois à travers la mise à jour de sa pondération. Il est donc nécessaire de corriger le poids 2017 par l'évolution de la valeur ajoutée entre 2015 et 2017.

Avec les notations de la section 1.2 et en notant va_i^{A-1} la part en VA de la branche i dans l'agrégat immédiatement supérieur (« Agg »), les pondérations corrigées s'écrivent alors (quel que soit le niveau de nomenclature) :

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

Comme il est nécessaire de connaître l'indice agrégé en moyenne annuelle de l'année précédente pour le calcul, il faut progresser de façon itérative en partant des niveaux les plus fins (approche montante « *bottom-up* ») :

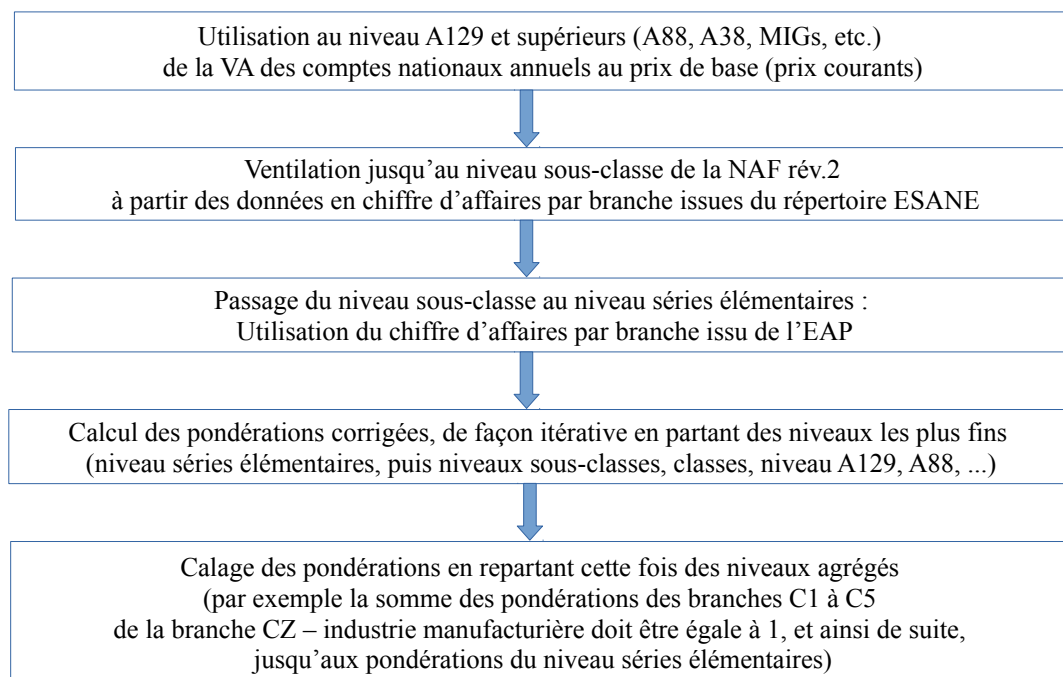
- on part des indices élémentaires qui sont connus (*cf.* chapitre précédent) et on calcule les indices moyens annuels au niveau élémentaire, ce qui permet de calculer les pondérations à ce niveau ;
- à l'aide des indices élémentaires et des pondérations, on peut calculer les indices au niveau supérieur ;
- de nouveau, on peut calculer les pondérations à l'aide des VA connues par ailleurs et les indices annuels moyens ;
- et ainsi de suite, jusqu'aux niveaux les plus agrégés.

Enfin, afin de respecter les propriétés d'additivité, on procède en sens inverse pour recalculer les poids (approche descendante « *top-down* ») afin que la somme des poids des sous-branches d'un agrégat soit égale à 1. L'ensemble des opérations pour le calcul des pondérations est résumé dans la figure 1.

33 À noter qu'une relation directe existait entre la NAF rev. 1 et la nomenclature CC.

Figure 1 : principales étapes du calcul des pondérations

2.2.4- La rétopolation des pondérations sur longue période



Au vu des nombreuses sources nécessaires pour le calcul des pondérations, celles-ci ont pu être calculées de façon standard à partir de 2010. Auparavant, les évolutions des nomenclatures ou des sources (par exemple le passage des EAB à l'EAP) n'ont pas permis un calcul systématique sous cette forme.

Afin de pouvoir chaîner les indices sur longue période (depuis 1990), il a été nécessaire de procéder à une rétopolation à partir des informations disponibles : comptes nationaux en branche disponibles sur longue période (niveau A129 depuis 1999 et niveau A88 avant), et aux niveaux plus fins, utilisation des pondérations des anciennes bases de l'IPI et interpolation entre deux années de référence. Avant 2000, la structure « *infra* NAF A88 » est ainsi constante.

2.2.5- Mise à jour des pondérations et révisions

Compte tenu des délais de disponibilité des données rentrant dans la construction des pondérations (le compte annuel dit « semi-définitif » pour l'année 2017 sera par exemple publié en mai 2019), les indices récents (année N) s'appuient, dans un premier temps tout au moins, sur des pondérations calculées sur des données antérieures, en attendant que les données N-1 (année de référence utilisée pour le calcul des pondérations) soient publiées. Ces opérations peuvent naturellement générer quelques révisions sur les dernières années.

3- Pour aller plus loin : interprétation des formules de chaînage pour les évolutions mensuelles et annuelles

En faisant apparaître les valeurs ajoutées dans la formule de chaînage, on obtient après simplification :

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

avec VA_i^{A-1} la valeur ajoutée de la branche i pour l'année $A-1$.

L'évolution de l'indice agrégé entre deux mois (m) et ($m+1$) d'une même année (m ne peut être le mois de décembre) s'écrit alors :

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

L'évolution correspond donc simplement à la moyenne des évolutions élémentaires, pondérées par la part des sous-branches dans l'agrégat. De la même façon, l'évolution annuelle entre ($A-1$) et (A) s'écrit

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

Autrement dit, pour calculer l'évolution annuelle, on part de la structure en valeur ajoutée de l'année précédente et on applique les évolutions des indices en moyenne annuelle.

La relation entre les mois de janvier de l'année A et de décembre de l'année $A-1$ est quant à elle plus complexe, puisqu'elle intègre implicitement le choc lié au changement de pondérations (passage des pondérations s'appuyant sur la VA $A-2$ aux pondérations s'appuyant sur la VA $A-1$ avec le changement d'année).