

Chapitre 5 -Le calcul des indices : le niveau élémentaire

Comme vu précédemment, l'IPI repose sur une enquête réalisée à un niveau fin dans le but d'obtenir des ensembles élémentaires relativement homogènes. Ce niveau élémentaire (non publié) correspond à la brique de base de l'IPI. Ce chapitre présente le mode de construction des indices élémentaires qui serviront ensuite au calcul des niveaux supérieurs (*cf.* chapitre 6).

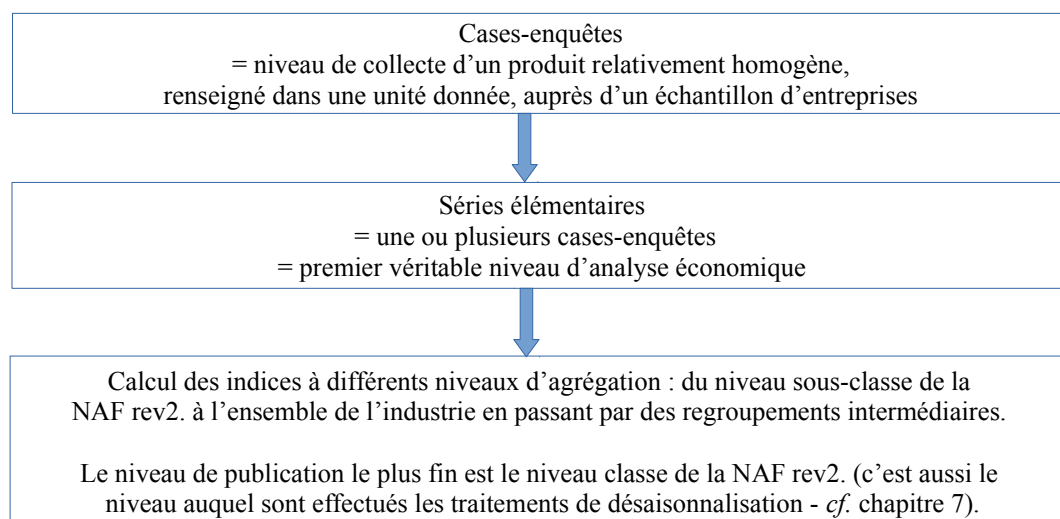
1- Collecte des données individuelles : le principe de la case-enquête

Comme présenté dans les chapitres précédents, les données individuelles des entreprises sont collectées sur des contours bien définis, appelés « cases-enquêtes », qui correspondent très précisément au croisement d'un produit (ProdEMB), d'une variable d'observation (quantités, facturations, heures travaillées) et d'une unité (tonnes, kilos, euros, etc.). Le terme de case-enquêtes se rapporte directement à la case du questionnaire où, pour un produit donné, l'entreprise (au sens de l'unité légale) remplit sa production pour le mois considéré. Le processus de collecte de l'IPI s'appuie sur un nombre important de cases-enquêtes. Par exemple, au sein de la NAF 2932Z – « Fabrication d'autres équipements automobiles », on dispose d'une case-enquête portant sur la « fabrication de sièges automobiles » avec comme variable d'observation les facturations, mais aussi une case-enquête sur les organes de direction et de freinage, une sur les parties et accessoires en matières thermoplastiques....

Cette nomenclature de collecte (l'ensemble des cases-enquêtes) peut être révisée en fonction des modifications apportées sur la nomenclature de produits européenne « ProdCom ». Elle est gérée dans un référentiel qui regroupe l'ensemble des produits observés dans l'EAP ²² (et qui sert de référence pour les EMB. Les produits observés dans les EMB (appelés ProdEMB) sont généralement des regroupements de plusieurs produits observés dans l'EAP (ou ProdEAP).

La construction de l'échantillon d'entreprises interrogées (environ 4 500 et 5 000 entreprises au total) est réalisée case-enquête par case-enquête. Les données sont collectées à travers le portail Coltrane de l'Insee de réponse aux enquêtes entreprises avant d'être contrôlées puis intégrées au calcul des indices.

Figure 1 : schéma général de construction des indices



²² Cf. chapitre 2 pour la présentation de cette source.

2- Méthode de calcul des indices élémentaires

2.1- Différents types de séries élémentaires

Différents modes de suivi

La construction de l'IPI commence par le calcul de séries élémentaires. Selon l'importance et la complexité des secteurs, ces séries élémentaires peuvent représenter un niveau de détail plus ou moins fin et peuvent utiliser différentes variables de mesure. Elles sont composées d'une ou plusieurs cases-enquêtes (*cf. supra*) homogènes (facturations et heures travaillées ne sont pas combinées par exemple). Les indices des séries primaires sont de moyenne annuelle 100 en année de référence.

L'IPI étant un indicateur de volume de production, les séries suivies en facturation doivent être déflatées (par un indice de prix) et les séries suivies en heures travaillées corrigées d'un coefficient technique prenant en compte l'évolution de la productivité dans l'activité étudiée.

Séries internes et externes

La majeure partie du suivi et de la collecte des données pour l'indice de la production industrielle est effectuée par l'Insee (séries internes). Néanmoins, dans certains cas, le calcul est directement effectué par un partenaire extérieur (série externe) : il s'agit notamment des séries de l'activité de l'industrie agroalimentaire ainsi que des séries des branches de l'énergie.

2.2- La définition d'un indice élémentaire

Un indice élémentaire correspond à un indicateur d'évolution calculé comme le rapport entre un numérateur et un dénominateur. Dans le cas des séries internes, le dénominateur est égal à la moyenne de l'année de base (la référence), la difficulté consiste donc à estimer un numérateur représentant l'évolution de l'activité par rapport à ce point de référence. Autrement dit, un indice élémentaire s'écrit :

$$Ind_i^{m,A} = \frac{Num_i^{m,A}}{\left(\sum_{k \in Aref} \frac{Num_i^{k,Aref}}{12} \right)} * 100$$

où :

- *i* désigne la branche suivie
- *m* et *A* respectivement le mois et l'année de calcul de l'indice
- *Aref* l'année de référence (ou année de base), actuellement 2015

Comme évoqué plus haut, dans le cas des séries suivies en facturations ou en heures travaillées, cet indice doit de plus être ajusté d'un coefficient technique prenant en compte l'évolution des prix ou de la productivité.

2.3- Calcul des numérateurs des séries élémentaires internes

Chaînage des séries élémentaires au mois le mois À partir d'un mois de référence, les numérateurs sont construits de proche en proche (chaînage) en appliquant les évolutions mensuelles calculées à partir des entreprises présentes deux mois consécutifs²³ (dénommées dans la suite « présentes-présentes »)²⁴.

Soient :

23 Ici consécutif ne signifie pas systématiquement deux mois qui se suivent dans le calendrier. Pour certains cas particuliers ou si une évolution entre deux mois qui se suivent n'est pas calculable, alors on chaînera par rapport au premier mois qui convient.

24 La méthodologie présentée ici est celle mise en œuvre depuis 2015. La méthodologie retenue précédemment était un peu différente et reposait sur le calcul d'un dénominateur commun à tous les mois d'une année, les numérateurs n'étaient donc pas directement chaînés de proche en proche.

Num_S^m = le numérateur de la série élémentaire « S » pour le mois m d'une année A ;

$EvolPP_S^{m/m-1}$ = l'évolution calculée sur l'ensemble des produits inclus dans la série « S » pour les entreprises présentes-présentes en m-1 et m (entreprises présentes dans l'échantillon pour les dates m et m-1 et pour lesquelles on dispose pour ces deux dates soit d'une réponse soit d'une imputation) ;

$M_{j \times case}^m$ = montant d'une unité légale « j » pour un produit correspondant à une case-enquête « case » de la série « S » (une entreprise donnée pouvant être interrogée pour différents produits au sein d'une série S) et $p_{j \times case}^m$ son poids de sondage ;

$(j, case) \in [S \times [m; m-1]]$ l'ensemble des couples *entreprise x case-enquête* présentes-présentes interrogées à la fois en m et m-1 pour les cases-enquêtes de la série S.

Alors, en calculant l'évolution pondérée entre m et m-1

$$EvolPP_S^{m/m-1} = \frac{\sum_{(j, case) \in [S \times [m; m-1]]} p_{j \times case}^m M_{j \times case}^m}{\sum_{(j, case) \in [S \times [m; m-1]]} p_{j \times case}^{m-1} M_{j \times case}^{m-1}},$$

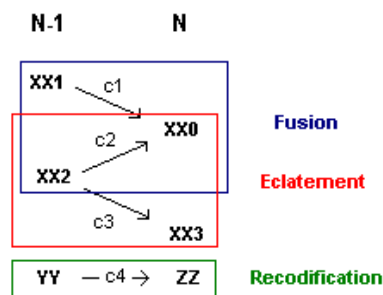
On peut calculer les numérateurs par chaînage de proche en proche :

$$Num_S^m = Num_S^{m-1} \times EvolPP_S^{m/m-1}$$

Au-delà de cette formule générale, un certain nombre de cas particuliers doivent être traités.

Cas particulier 1 : calcul de l'évolution entre décembre et janvier et gestion des cases-enquêtes Chaque année, les cases-enquêtes peuvent être modifiées (regroupements, recodification, éclatements).

Figure 2 : modifications possibles de cases-enquêtes



Le chaînage des montants déclarés par les entreprises d'un mois sur l'autre au moment du changement d'année peut poser difficulté, car les déclarations ne sont pas comparables sur des cases-enquêtes différentes. Il est alors nécessaire d'estimer des coefficients de passage permettant de passer (dans les cas où il y a eu un changement) des cases-enquêtes d'une année à celles de l'année suivante.

Cas particulier 2 : chaînage par rapport aux mois atypiques de juillet et d'août

Les mois de juillet et d'août étant des mois atypiques, il n'est pas souhaitable de chaîner des évolutions en les prenant pour référence. En conséquence, de façon générale pour les méthodes utilisées pour l'IPI, lors du calcul du mois d'août et de septembre, la comparaison se fait par rapport au mois de juin (dans les formules de calcul *supra*, il suffit alors de remplacer « m-1 » par juin).

Cas particuliers 3 : gestion des « 0 »

Lors du calcul de $EvolPP_S^{m/m-1}$, il est possible que le dénominateur soit nul. Dans ce cas, on chaînera par rapport au premier mois observé non nul (m-2, m-3, etc.). De même si le numérateur par rapport auquel on chaîne est nul.

3- Imputation des non-réponses

La méthode d'imputation d'une donnée individuelle j (entreprise dans une case-enquête C) repose sur la combinaison des quantités suivante :

- le ratio m/m-1 du total de la case-enquête (ensemble des entreprises répondantes pour cette case-enquête en m et présentes en m-1 (répondante ou imputée)) : $EvolR_C^{m/m-1}$;
- le ratio m-12/m-13 de la série élémentaire S dans laquelle se trouve la case-enquête : $EvolSerie_S^{m-12/m-13}$;
- le tout pondéré par un taux de réponse dans la case- enquête qui permet de prendre en compte la « qualité » de l'évolution : TxR_C^m .

$$M_j^m = M_j^{m-1} \times [(EvolR_C^{m/m-1})^{TxR_C^m} \cdot (EvolSerie_S^{m-12/m-13})^{(1-TxR_C^m)}]$$

Cas particulier 1 : gestion des 0

Si dans la formule ci-dessus un des deux ratios est nul (ou si aucune entreprise ne répond) : alors la formule devient :

$M_{i,ce(an_m)}^m = M_{i,ce(an_m)}^{m-1} \times Evol$ avec $Evol$ le ratio non nul ou calculable (dans le cas d'un taux de réponse nul). Si les deux évolutions sont nulles, le montant imputé vaut 0.

Cas particulier 2 : entreprises entrantes et non-répondantes :

Les entreprises qui sont nouvelles dans l'échantillon et non-répondantes ne peuvent être imputées. Elles ne sont donc pas utilisées pour le calcul le premier mois.

4- Gestion des mouvements infra-annuels dans l'échantillon (corrections de chaînage)

L'échantillon d'entreprises, tiré une fois par an, peut connaître des modifications tous les mois, notamment en raison des restructurations (rachat d'une entreprise par une autre, séparation...). Ainsi, de nouvelles entreprises peuvent être créées, intégrant une ou plusieurs entreprises qui étaient interrogées les mois précédents. Des entreprises peuvent également cesser leur activité ou encore rester mais sur un champ d'activité complètement différent.

Ces modifications d'échantillons (modifications prises en compte *via* l'application de gestion des enquêtes mensuelles de branches) interfèrent avec le calcul des séries élémentaires. En effet le calcul des indices étant chaîné entre les mois m-1 et m, plusieurs modifications de contours peuvent l'affecter :

- pour les entreprises restructurées qui ne changent pas d'identifiant, mais qui connaissent des modifications de champ ou de comportement de réponses, on souhaite disposer de montants comparables entre m et m-1 ;

- pour les nouvelles entreprises restructurées intégrant une ou plusieurs entreprises de l'échantillon, on veut également disposer de montants comparables entre m et $m-1$, même si toutes les entreprises du nouveau contour de l'entreprise restructurée n'étaient pas interrogées en $m-1$;
- les suppressions en revanche n'interviennent pas dans le chaînage et donc dans le calcul d'indice²⁵.

Dans tous ces cas, il est nécessaire de mettre en place des corrections de chaînage qui permettront de conserver des évolutions pertinentes. Ces corrections sont mises en place en concertation avec les entreprises quand cela est possible.

Soit :

Y_{m-1} : montant de correction de chaînage pour un croisement (entreprise(siren) x case-enquête) pour le mois $m-1$ et comparable en matière de champ à une déclaration en m , notée X_m ;

X_m : montant déclaré pour un croisement (entreprise(siren) x case-enquête) en m (réponse de l'entreprise ou imputation).

Cas 1 : création d'entreprise

Dans le cas d'une création d'entreprise, on introduit une correction en $m-1$ pour pouvoir calculer l'évolution à champ comparable : le montant de correction de chaînage est utilisé pour le calcul du numérateur du mois m (chaînage entre m et $m-1$). Par contre, il ne doit pas être utilisé dans le calcul de $m-1$ (il n'existe pas de réponse en $m-2$).

m-2	m-1	m
	Y_{m-1}	X_m

 chaînage d'un mois à l'autre

Si une entreprise est créée, mais qu'aucun montant de correction de chaînage Y_{m-1} n'a été saisi (par exemple parce qu'il n'est pas possible d'avoir une estimation fiable de ce montant), l'entreprise n'est pas prise en compte lors du calcul du mois m . Elle sera alors intégrée pour la première fois dans le calcul du mois suivant.

Cas 2 : restructuration ou modification de comportement de réponse

Dans le cas d'une modification d'une entreprise (restructuration ou changement de comportement de réponse), le montant de correction de chaînage est utilisé pour le calcul du numérateur du mois m . En revanche, dans le calcul de $m-1$, le montant utilisé est toujours celui du mois précédent, saisi par l'entreprise (ou modifié par le gestionnaire, ou imputé).

m-2	m-1	m
X_{m-2}	X_{m-1}	X_m
	Y_{m-1}	

 chaînage d'un mois à l'autre

²⁵ Généralement, si ce n'est pas déjà le cas, le produit concerné est mis à 0 avant retrait effectif de l'échantillon de calcul afin de prendre en compte cette suppression.

Si une entreprise est restructurée ou modifie son comportement de réponse, mais qu'aucun montant de correction de chaînage n'a été saisi, le montant utilisé est X_{m-1} .

5- Priorisation de l'analyse des réponses individuelles au sein d'une série élémentaire

Les séries élémentaires sont expertisées par des gestionnaires en direction régionale, afin de détecter d'éventuelles réponses erronées ou des évolutions atypiques. Compte tenu des délais de production très contraints, il n'est pas possible de vérifier toutes les données. L'analyse des séries est alors priorisée selon deux critères principaux :

- les contributions individuelles au glissement annuel (entre $m/m-12$) de la série élémentaire ;
- les contributions individuelles à la révision de l'évolution entre $m-1$ et m de la série élémentaire.

En reprenant les notations déjà utilisées et si $\{j \in [S \times (m - 12)]\}$, la contribution d'une entreprise i (présente en m et $m-12$) au glissement annuel « $m/m-12$ » d'une série élémentaire s'écrit :

$$\text{Contribution}_{i \rightarrow S}^{m/m-12} = \frac{p_i^m * M_i^m - p_i^m * M_i^{m-12}}{\sum_{j \in [S \times (m-12)]} p_j^m * M_j^{m-12}}$$

Par ailleurs, en notant :

- « dca » la date de campagne actuelle et « dca-1 » la date de campagne précédente, c'est-à-dire pour les indices calculés le mois précédent ;
- $M_i^{m;dca}$ le montant déclaré par l'entreprise i pour le mois m , lors de la campagne de calcul « dca » ;
- $\{j \in [S \times (m; m-1); dca]\}$ l'ensemble des (entreprises x cases enquêtes) présentes-présentes en m et $m-1$ pour la série S lors de la campagne mensuelle « dca » ;
- $\{j \in [S \times (m; m-1); dca-1]\}$ l'ensemble des (entreprises x cases enquêtes) présentes-présentes en m et $m-1$ pour la série S lors de la campagne mensuelle précédente.

Alors la contribution d'une entreprise i à la révision de l'évolution mensuelle « $m/m-1$ » d'une série élémentaire s'écrit :

$$\text{Contribution}_{i \rightarrow S}^{\text{Révision } m/m-1, \text{ campagne dca}} = \frac{p_i^m * M_i^{m;dca}}{\sum_{j \in [S \times [m, m-1]; dca]} p_j^m * M_j^{m-1; dca}} - \frac{p_i^m * M_i^{m;dca}}{\sum_{j \in [S \times [m, m-1]; dca-1]} p_j^m * M_j^{m-1; dca}}$$