

Calcul d'un indice des prix des produits de grande consommation dans la grande distribution

Note méthodologique

Janvier 2014

CONTEXTE DE LA PRÉSENTE ÉTUDE : LES MODIFICATIONS INTERVENUES EN JANVIER 2014 DANS LA PUBLICATION DE L'INDICE DES PRIX DE LA GRANDE DISTRIBUTION 2

1 DÉFINITIONS ET POSITION DU PROBLÈME : LE CHOIX DU NIVEAU ÉLÉMENTAIRE D'AGRÉGATION DANS UN INDICE DES PRIX 6

1.a Généralités sur le calcul d'un indice des prix 6

1.a.1 Le concept de panier-type 6

1.a.2 La méthode de calcul de l'IPC 6

1.a.3 Représentation schématique de la méthode de calcul IPC 7

1.b Notre objet : Un indice des prix restreint à un sous-champ de la consommation 8

1.b.1 Les produits de grande consommation 8

1.b.2 Les formes de vente 9

1.c Impact sur le niveau élémentaire d'agrégation de la restriction à un sous-champ 10

1.c.1 Le problème du découpage des cellules élémentaires 10

1.c.2 La méthode de calcul des IGD jusqu'en décembre 2013 10

1.c.3 Écarts d'indice induits par la différence de méthode 10

1.c.4 Discussion théorique sur la méthode IGD-2004 12

2 RECHERCHE DU « MEILLEUR » NIVEAU ÉLÉMENTAIRE D'AGRÉGATION 13

2.a Pourquoi modifier le niveau élémentaire d'agrégation ? 13

2.b Un nouveau découpage du territoire 14

2.c Les indices par forme de vente obtenus en utilisant le nouveau découpage géographique 16

Contexte de la présente étude :

Les modifications intervenues en janvier 2014 dans la publication de l'indice des prix de la grande distribution

L'Insee publie depuis 2005, au format « Informations Rapides » et dans la Banque de Données Macroéconomiques (BDM, série G421), onze séries d'indices de prix à la consommation se rapportant aux produits dits « de grande consommation » et établissant des comparaisons entre l'inflation subie par ces produits dans la grande distribution¹ et dans les autres formes de vente.

La publication de ces indicateurs par l'Insee est prévue par l'article 7 du texte d'« engagement pour une baisse durable des prix à la consommation » signé le 17 juin 2004 par vingt-deux représentants de grands groupes commerciaux et fédérations d'entreprises en présence du Ministre de l'Économie, des Finances et de l'Industrie et du Ministre délégué aux petites et moyennes entreprises, au commerce, à l'artisanat, aux professions libérales et à la consommation.

Afin d'éclairer au mieux le débat économique, l'indice des biens de grande consommation dans la grande distribution est complété par cinq séries d'indices déclinant l'ensemble des biens de grande consommation en grands groupes de produits (viandes, boissons, autres produits alimentaires hors produits frais, ensemble des produits alimentaires hors produits frais², et produits de grande consommation non alimentaires), ainsi que par cinq séries d'indices portant sur le même champ des produits de grande consommation mais se rapportant à différents groupements de formes de vente (hypermarchés, supermarchés, grande distribution étendue³, formes de vente hors grande distribution étendue, et toutes formes de vente).

En complément, la publication au format « Informations Rapides » trace en première page un graphique comparant les évolutions en glissement annuel de deux de ces indices : l'indice des biens de grande consommation dans la grande distribution d'une part et l'indice des biens de grande consommation dans les formes de vente hors grande distribution étendue (dites « autres formes de vente ») d'autre part.

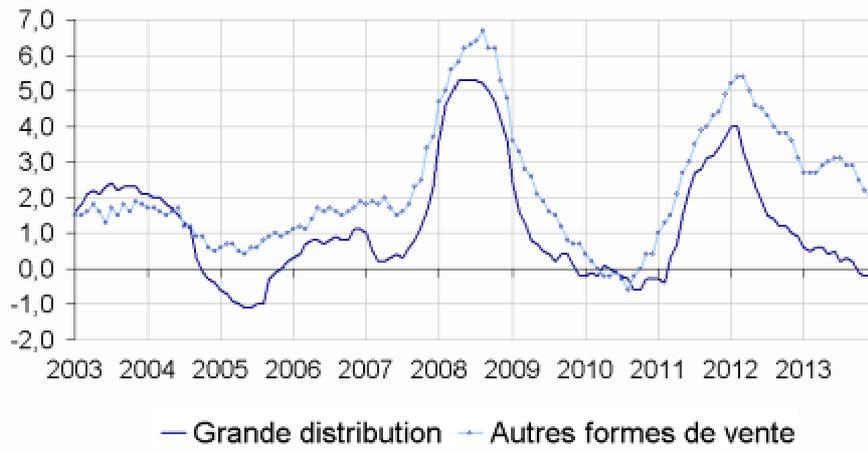
Graphique 0.1 : extrait de la publication de l'indice des prix dans la grande distribution – décembre 2013

¹ La grande distribution est définie comme l'ensemble des hypermarchés et supermarchés (hors hard discount), c'est-à-dire comme l'ensemble des entreprises dont l'activité principale est le « commerce de détail en magasin non spécialisé à prédominance alimentaire » et dont la surface de vente est au moins égale à 400 m². Pour une définition précise des hypermarchés et supermarchés, on se reportera aux sections 47.11D et 47.11F de la Nomenclature d'Activité Française (NAF rév.2), en ligne sur le site insee.fr, rubrique Définitions et méthodes, section Nomenclatures : http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/naf2008/n4_47.11.htm

² Puisque les produits frais sont exclus du texte d'engagement précité.

³ Hypermarchés, supermarchés, supérettes et magasins populaires.

Variations annuelles en %



Les évolutions en glissement annuel permettent d'apprécier, pour un mois donné, l'inflation subie par les biens de grande consommation dans la forme de vente considérée entre le mois en question et le même mois de l'année précédente. Cela permet d'avoir une vision de l'inflation à plus long terme et indépendamment des phénomènes saisonniers.

Lorsque l'évolution en glissement annuel est négative, cela signifie que les prix ont baissé depuis le même mois de l'année d'avant. C'est le cas par exemple dans la grande distribution entre décembre 2012 et décembre 2013 : les prix ont baissé de 0,2 %. Dans le même temps, ils ont augmenté de 2,0 % dans les formes de vente hors grande distribution étendue.

Le fait que la courbe « grande distribution » soit systématiquement située sous la courbe « autres formes de vente » depuis fin 2004 signifie que l'inflation observée sur les produits de grande consommation est toujours plus faible dans la grande distribution que dans les formes de vente hors grande distribution étendue. Lorsque l'écart entre les courbes s'accroît, cela signifie que le différentiel d'inflation entre les formes de vente s'accroît. C'est le cas entre mi-2011 et mi-2013, où le différentiel s'accroît de 2 points en deux ans.

**Tableau 0.1 : Variations annuelles des indices selon la forme de vente
– Produits de grande consommation**

		Grande distribution	Autres formes de vente	Différentiel
2011	Janvier	-0,3 %	+0,9 %	+1,2 %
	– Juillet	+2,7 %	+3,5 %	+0,8 %
2012	Janvier	+3,9 %	+5,2 %	+1,3 %
	– Juillet	+1,5 %	+4,3 %	+2,8 %
2013	Janvier	+0,7 %	+2,7 %	+2,0 %
	– Juillet	+0,2 %	+3,1 %	+2,9 %

On remarque que le différentiel se creuse lors de la décélération des prix de 2012, qui se manifeste moins rapidement dans les formes de vente hors grande distribution étendue que dans les formes de vente de la grande distribution. De plus, il y a à nouveau une accélération des prix dans les autres formes de vente au cours du premier semestre 2013, alors que la décélération se poursuit toujours dans la grande distribution.

Ce constat, dont l'interprétation économique n'est pas immédiate, nous a poussé à faire un audit de la méthode de calcul de ces indices et à en proposer des améliorations. Ces améliorations, qui entrent en vigueur dès la publication de janvier 2014, sont exposées dans la partie 1 du présent document.

Par ailleurs, puisque les améliorations apportées à la méthode ne modifient pas en profondeur l'allure des courbes d'indices, on propose en partie 2 quelques analyses statistiques complémentaires qui fournissent des pistes pour la compréhension des différences durables d'inflation qui sont observées entre la grande distribution et les autres formes de vente sur le champ des produits de grande consommation.

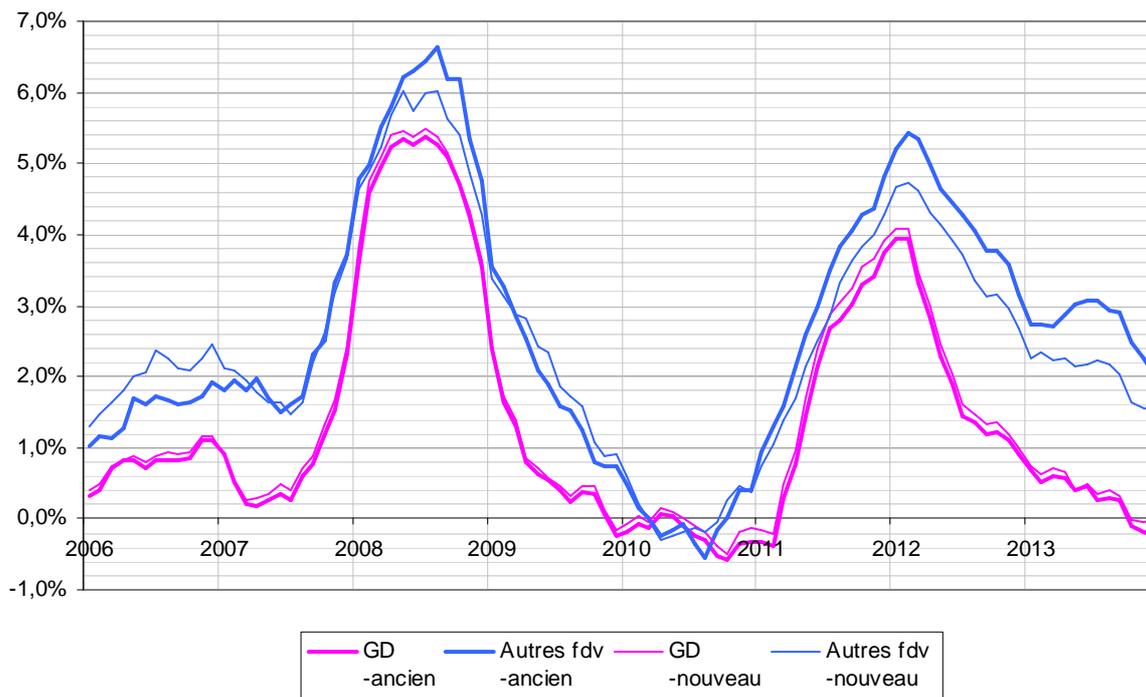
Ces exercices se heurtent aux limites de la méthode : le nombre d'observations à notre disposition est en effet trop faible pour que l'on puisse entrer dans un niveau de détail suffisamment fin pour mener à terme les analyses. Pour la même raison, l'Insee a jugé à l'issue de ces travaux que les indices portant sur les sous-champs « hypermarchés », « supermarchés » et « formes de vente hors grande distribution étendue » étaient insuffisamment robustes pour être exploités. Pour cette raison, ces indices ne sont plus publiés à partir de janvier 2014. La courbe nommée « autres formes de vente » dans la publication « Informations rapides » correspond dorénavant au champ des formes de vente hors grande distribution. Par rapport à l'ancienne série, la nouvelle série « autres formes de vente » inclut donc les supérettes et les magasins populaires.

La méthode de calcul ayant changé, la publication de la série G421 dans la BDM s'arrête en décembre 2013. Nous invitons les utilisateurs de ces indices à se reporter sur la nouvelle série BDM créée à partir de janvier 2014 et qui contient les nouvelles séries d'indices liées à la publication « indice des prix dans la grande distribution », présentées en séries longues depuis 1998.

Le changement de méthode n'a que très peu d'impact sur les évolutions en glissement annuel. Dans la nouvelle publication, le graphique de la première page est donc très semblable à l'ancien graphique. L'impact le plus important sur ce graphique provient du changement de définition pour la courbe « autres formes de vente ». Ci-dessous, on superpose l'ancien et le nouveau graphique, de manière à se faire une idée des modifications apportées.

**Graphique 0.2 : Evolutions en glissement annuel
– Produits de grande consommation**

Les courbes en traits pleins correspondent à l'ancienne publication, ceux en traits fins à la nouvelle. L'écart entre les deux courbes roses est un pur artéfact méthodologique : on peut constater qu'il est très faible. L'écart entre les deux courbes bleues tient compte à la fois du changement de méthode et du changement de définition : c'est pourquoi il est plus important.

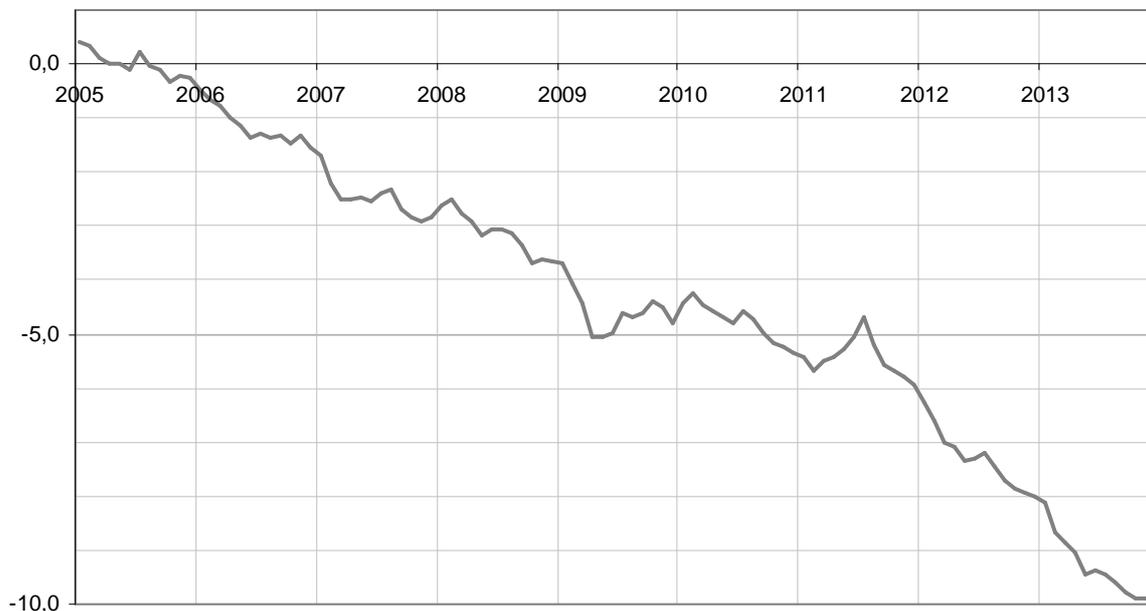


Les observations faites à partir des anciennes définitions restent inchangées avec les nouvelles :

- L'inflation est moins forte dans la grande distribution,
- La période de déflation de 2010 a été plus longue dans la grande distribution,
- Les pics d'inflation ont lieu aux mêmes périodes (pas d'effet retard d'une forme de vente sur une autre) mais dans des amplitudes différentes,
- Le différentiel d'inflation entre les formes de vente se creuse entre mi-2011 et mi-2013.

Puisque ces tendances s'établissent sur le long terme, les écarts d'indice entre formes de vente se cumulent : les indices dans la grande distribution et dans les autres formes de vente divergent sur longue période, ainsi que le montre le graphique en page suivante.

Graphique 0.3 : Écart de l'indice d'ensemble des produits de grande consommation dans la grande distribution à l'indice d'ensemble des produits de grande consommation dans les autres formes de vente, les deux indices considérés en base 100 sur l'année 2005



Au début de la période d'observation, l'écart d'inflation s'établissait en faveur des formes de vente hors grande distribution. En moyenne sur 2005, l'inflation a été la même en et hors grande distribution. Mais depuis, l'écart se creuse tendanciellement. En neuf ans, l'indice des prix des produits de grande consommation dans la grande distribution et le même indice considéré hors grande distribution ont cumulé un écart de près de 10 points d'indice.

On note une accélération de la divergence depuis mi-2011 : en seulement 2 ans et demie, les indices se sont écartés de 5 points supplémentaires – autant que durant les six années précédentes.

C'est pour tenter de comprendre cet écart, son creusement au fil des années et son accélération récente que la présente étude a été menée durant l'été 2013. Dans la suite du document, les tableaux et graphiques ne comportent pas de chiffres postérieurs à mai 2013, la publication de mai étant la dernière disponible au moment de l'étude.

1 Définitions et position du problème : le choix du niveau élémentaire d'agrégation dans un indice des prix

1.a Généralités sur le calcul d'un indice des prix

1.a.1 LE CONCEPT DE PANIER-TYPE

L'indice des prix à la consommation (IPC) est un indice de panier-type. Cela signifie que son calcul est basé sur la définition préalable d'un « panier » de biens et services supposé « représentatif » de la consommation d'ensemble sur le territoire considéré. Cette notion de représentativité se traduit dans les faits par un calage des échantillons sur un certain nombre de dimensions :

- Dans la définition des produits suivis (dimension sectorielle) : on utilise la nomenclature de produits Coicop, que l'on décline ensuite en postes. L'ensemble des 305 postes de l'IPC couvre en théorie l'ensemble du champ de la consommation en France. En pratique, le taux de couverture de l'IPC est de 97% depuis son rebasage en 1998. Dans chacun de ces 305 postes, on définit ensuite une liste de produits-types, aussi appelés variétés de l'IPC. Les variétés sont des « représentants » du poste et présentent des définitions précises de manière à ce que les relevés de prix se rapportant à une même variété soient comparables entre eux. Ces relevés peuvent toutefois se rapporter à des situations d'achat très différentes, en fonction des dimensions suivantes.
- Dans la définition des agglomérations (dimension géographique) : la liste des agglomérations de l'IPC est issue d'un plan de sondage stratifié en fonction de la taille démographique de l'agglomération. L'IPC est collecté dans une centaine d'agglomérations en métropole et dans les DOMs.
- Dans la définition des points de vente (dimension activité du vendeur) : la liste des produits du panier-type respecte une répartition par type de points de vente qui est conforme aux parts de marché détenues par ce type de points de vente, pour la variété considérée.

Définir le panier-type de l'IPC revient donc à définir une liste de produits à l'intérieur d'une liste de triplets (variété de produits ; agglomération ; forme de vente), étant données des cibles de nombre de relevés à faire à l'intérieur de chaque triplet. Ces cibles d'effectifs sont obtenues à partir de données sur la consommation, dont la principale source est la comptabilité nationale.

Si l'échantillon des agglomérations de l'IPC est maintenue constante sur de longues périodes, l'échantillon des variétés de produits est révisée à la marge chaque année, de même que la liste des produits à l'intérieur de chaque triplet (variété de produits ; agglomération ; forme de vente).

La liste des produits est définie en décembre puis maintenue constante toute l'année, de manière à pouvoir calculer chaque mois un indice des prix mesurant l'évolution des prix depuis le mois de décembre de l'année précédente. On publie ensuite un indice « chaîné » depuis 1998, c'est-à-dire dont la valeur représente l'évolution des prix depuis l'année 1998, de manière à permettre de tracer les évolutions de prix sur le long terme⁴.

1.a.2 LA MÉTHODE DE CALCUL DE L'IPC

L'IPC utilise une formule d'agrégation dite de Laspeyres. Concrètement, l'indice à chaque niveau de la nomenclature des produits est le résultat de la moyenne arithmétique des indices du niveau de nomenclature inférieur, pondérée par les valeurs de consommation de la période de référence. Au niveau « variété », cela se traduit par la formule suivante :

Formule 1.a.1 : Méthode de calcul de l'IPC

– Agrégation par variété

$$I_{m,a}(v) = \sum_{\forall ag} \pi_a^v(ag) I_{m,a}(v, ag)$$

⁴ Pour plus de détails sur la méthodologie de l'indice des prix à la consommation, on pourra se reporter à l'ouvrage *Pour comprendre l'indice des prix*, dans la collection « Insee Méthodes », 1998.

où

$I_{m,a}(v)$ est l'indice de la variété v , pour le mois m et l'année a ,

$I_{m,a}(v, ag)$ est l'indice du couple variété v x agglomération ag , pour le mois m et l'année a ,

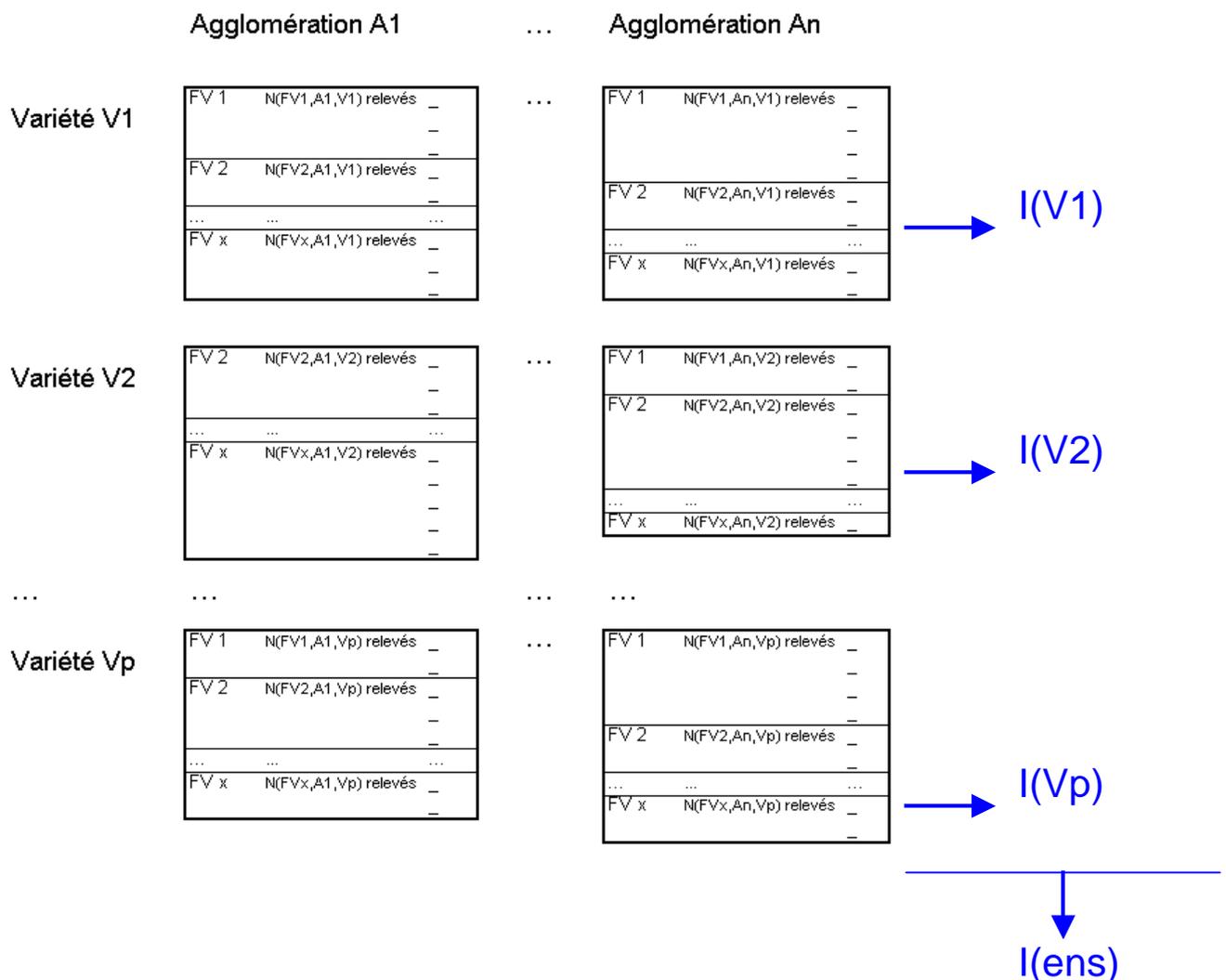
$\pi_a^v(ag)$ est la part de dépense en valeur de l'agglomération ag , pour la variété v et l'année a ,

avec $\forall v, \sum_{\forall ag} \pi_a^v(ag) = 1$

Les indices $I_{m,a}(v, ag)$ sont les indices de plus bas niveau calculés dans l'IPC. Ils ne peuvent pas être calculés par la formule de Laspeyres, car on ne dispose pas de données utilisables pour pondérer les observations élémentaires à l'intérieur de chaque couple (variété ; agglomération). Ces indices de premier niveau (ou niveau élémentaire) sont donc calculés à l'aide de formules d'agrégation non pondérées. En accord avec la théorie des indices, deux formules sont utilisées alternativement selon le type de variété :

- Si la variété est « homogène », c'est-à-dire que les produits qui la composent sont très comparables les uns aux autres en termes de qualité, alors on utilisera une formule de Dutot, qui consiste à faire le rapport des prix moyens arithmétiques entre la période courante et la période de référence ;
- Sinon, on utilisera une formule de Jevons, qui consiste à faire le rapport des prix moyens géométriques entre la période courante et la période de référence.

1.a.3 REPRÉSENTATION SCHEMATIQUE DE LA METHODE DE CALCUL IPC



1.b Notre objet : Un indice des prix restreint à un sous-champ de la consommation

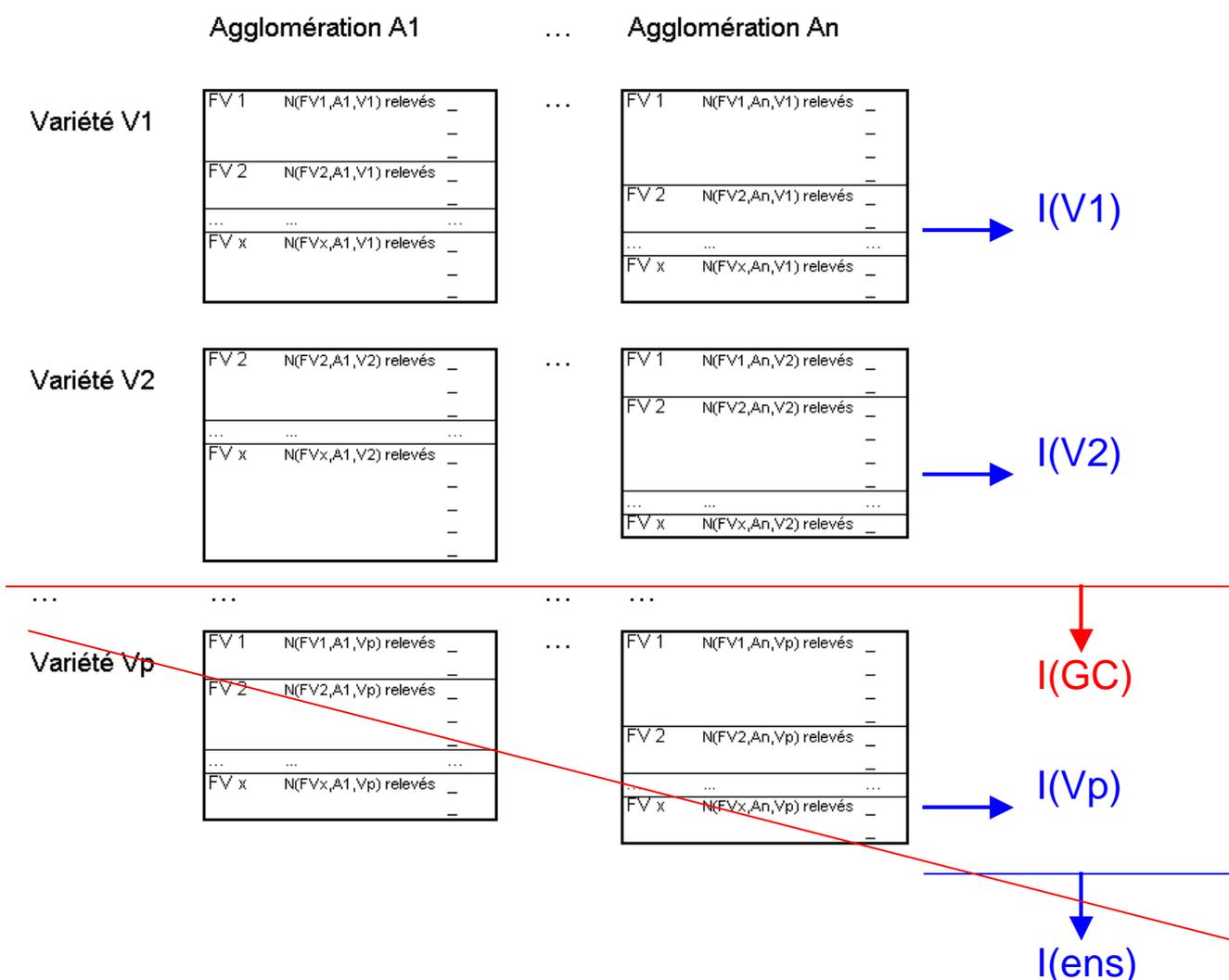
1.b.1 LES PRODUITS DE GRANDE CONSOMMATION

À l'intérieur du panier-type défini annuellement pour l'IPC, on souhaite restreindre notre champ d'intérêt aux seuls produits dits « de grande consommation », comprenant les groupes suivants de la nomenclature Coicop :

- + 01 Produits alimentaires et boissons non alcoolisées
- + 021 Boissons alcoolisées
- + 0561 Articles de ménage non durables
- + 1213 Appareils, autres articles et produits pour soins personnels

Les formules d'agrégation ne sont pas modifiées par cette restriction. L'indice des produits de grande consommation est simplement un sous-indice de l'IPC.

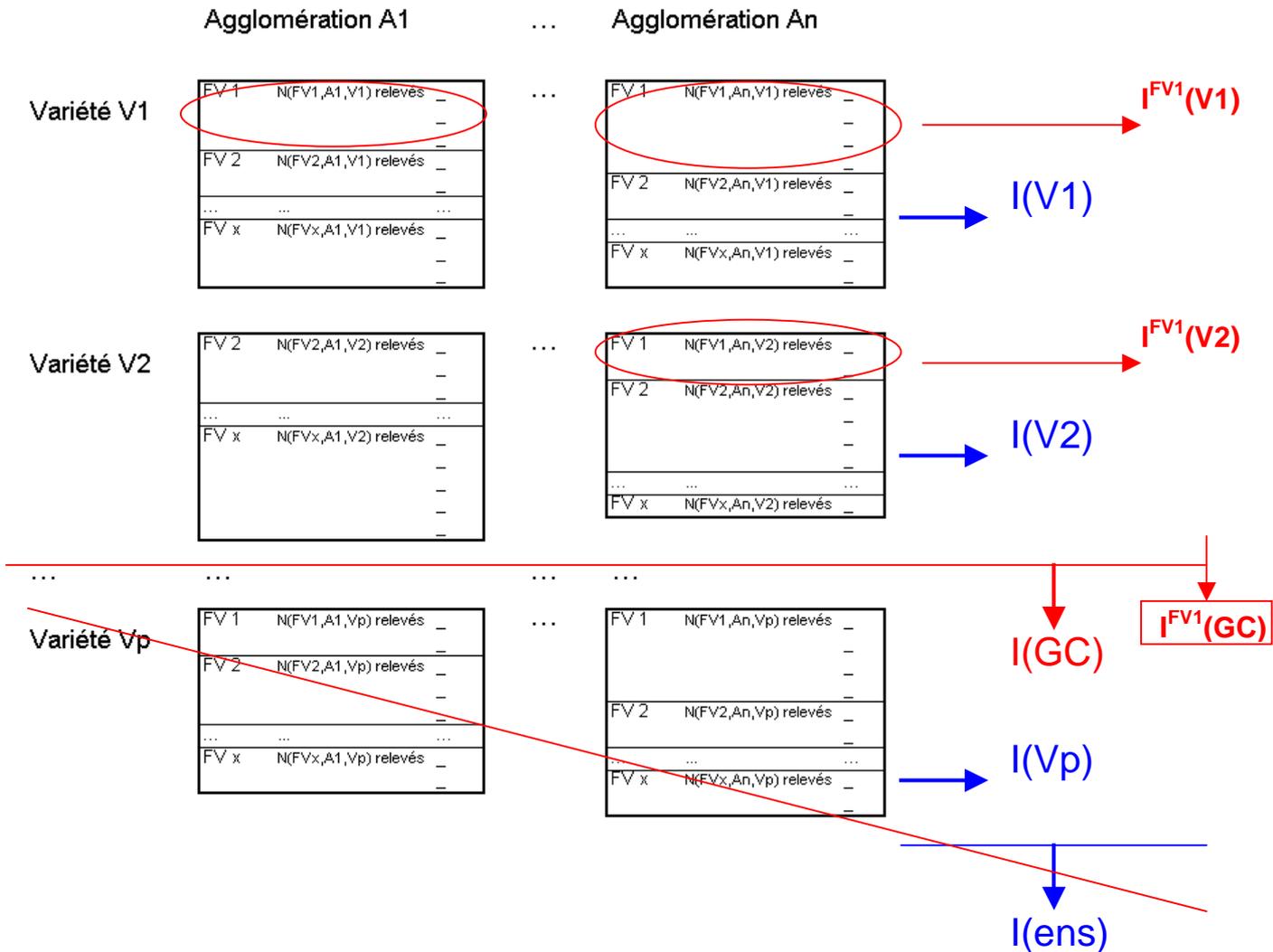
Schéma 1.a.1 : L'agrégation sur un sous-champ sectoriel



1.b.2 LES FORMES DE VENTE

La situation est différente lorsqu'on souhaite également restreindre le champ sur une partie de l'échantillon des points de vente, car le calcul des indices de niveau élémentaire est alors modifié.

Schéma 1.a.2 : L'agrégation sur un sous-champ sectoriel, avec sélection de certaines formes de vente au niveau élémentaire



Dans le cas qui nous intéresse, les formes de vente sont regroupées en deux grandes familles :

- Grande distribution : hypermarchés et supermarchés, c'est-à-dire les points de vente à dominance alimentaire de surface supérieure ou égale à 400 m². Sont exclus les points de vente de hard discount.
- Autres formes de vente : tous les points de vente qui ne sont ni hypermarché ni supermarché ; cela recouvre : les grandes surfaces spécialisées, grands magasins, magasins populaires, supérettes, petits magasins traditionnels, marchés, ainsi que les points de vente de service, à quoi l'on ajoute les points de vente de hard discount.

1.c Impact sur le niveau élémentaire d'agrégation de la restriction à un sous-champ

1.c.1 LE PROBLÈME DU DÉCOUPAGE DES CELLULES ÉLÉMENTAIRES

Il est tout à fait possible de calculer un indice des prix des produits de grande consommation dans la grande distribution (respectivement, dans les autres formes de vente) en appliquant la méthode IPC habituelle, selon la formule 1.a.1. Toutefois, appliquer la méthode IPC telle quelle sur des champs d'observations restreints à une forme de vente peut conduire à des indices de mauvaise qualité statistique, en raison de la présence d'un grand nombre de cellules élémentaires trop petites. En d'autres termes, les échantillons de produits par (variété x agglomération) de l'IPC ne sont pas assez importants pour supporter d'être divisés.

En effet, l'échantillon de l'IPC a été optimisé de manière à assurer une précision suffisante au niveau global tout en limitant les coûts de collecte. Ainsi, toutes les variétés ne sont pas représentées dans toutes les agglomérations, mais lorsqu'elles le sont, on assure en général une taille d'échantillon d'au moins 4 relevés par cellule élémentaire (variété x agglomération).

Outre la dégradation de la précision de l'indice, travailler sur des cellules élémentaires trop petites peut également conduire à un biais de l'indice, lorsque la formule d'agrégation retenue au niveau élémentaire est une formule de Jevons, car celle-ci n'est pas une fonction linéaire des prix.

Or le filtrage des observations par forme de vente conduit à des cellules élémentaires trop petites dans la grande majorité des cas : lorsqu'on s'intéresse aux variétés de produits de grande consommation, on constate que 60 % des cellules (variété x agglomération) de la grande distribution comportent strictement moins de 4 relevés ; ce chiffre grimpe à 97 % pour l'ensemble des autres formes de vente, ce qui signifie que l'indice des produits de grande consommation dans les formes de vente hors grande distribution est construit à partir de cellules élémentaires très petites.

Il ne semble donc pas souhaitable d'appliquer simplement la méthode IPC pour le calcul des indices déclinés par forme de vente.

1.c.2 LA MÉTHODE DE CALCUL DES IGD JUSQU'EN DÉCEMBRE 2013

Compte tenu des remarques précédentes, il a été décidé lors de la création de l'indice de la grande distribution (IGD) en 2004, d'éviter le niveau d'agrégation par (variété x agglomération) et de faire du niveau variété le premier niveau d'agrégation. De cette manière, la taille des cellules élémentaires est suffisante, même en découpant le panier par forme de vente.

Les mêmes raisons méthodologiques (les échantillons devenant trop petits après restriction du champ) ont conduit à ne pas publier un indice spécifique au hard discount, et à renoncer à la publication de l'IGD dans les départements d'outre-mer.

Au-delà de la modification d'échelle de calcul des indices élémentaires, une formule unique (proche d'une formule de Dutot) a été adoptée quelle que soit la variété considérée.

Ce changement de méthode modifiait la formule de calcul des indices au niveau variété. Les indices de variété pour l'IGD différaient donc des indices de variété pour l'IPC d'ensemble.

Les séries d'indices agrégés par groupes de produits et ceux agrégés par forme de vente pour l'ensemble des produits de grande consommation étaient ensuite obtenues par formule de Laspeyres sur la base de ces indices élémentaires de variété.

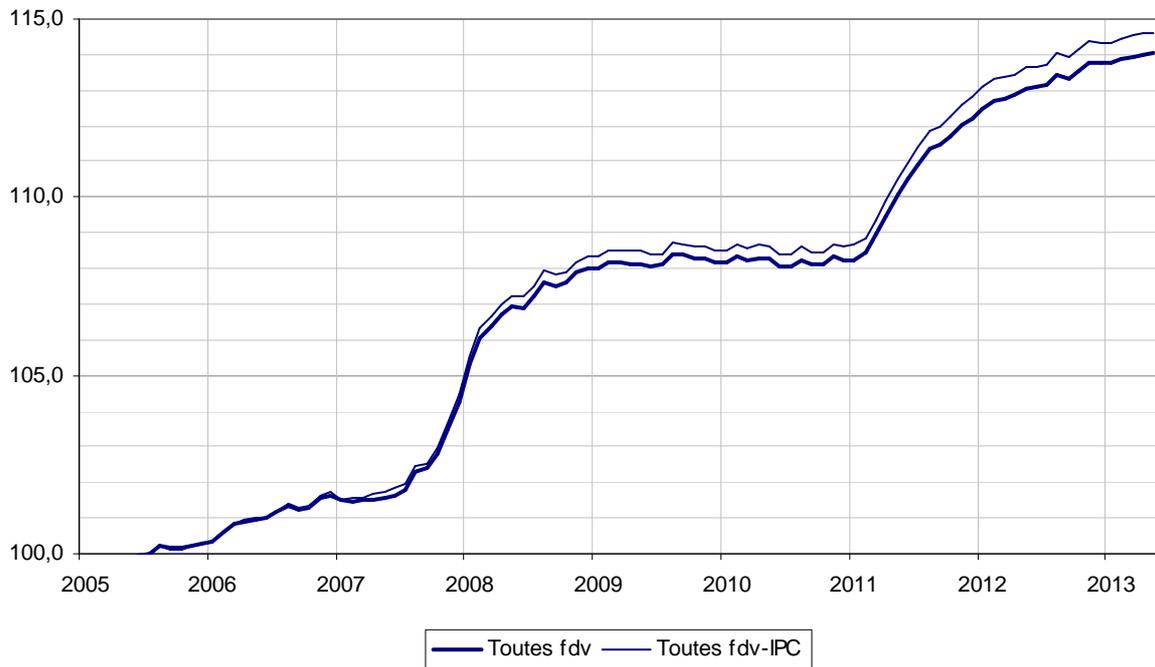
Dans la suite, on appellera cette méthode « méthode IGD-2004 », pour la distinguer de la nouvelle méthode en vigueur depuis janvier 2014, que l'on appellera « méthode IGD-2014 ».

1.c.3 ÉCARTS D'INDICE INDUITS PAR LA DIFFÉRENCE DE MÉTHODE

On cherche d'abord à déterminer si la méthode employée pour le calcul de ces indices présente une dérive globale par rapport à la méthode employée pour le calcul des IPC « classiques ».

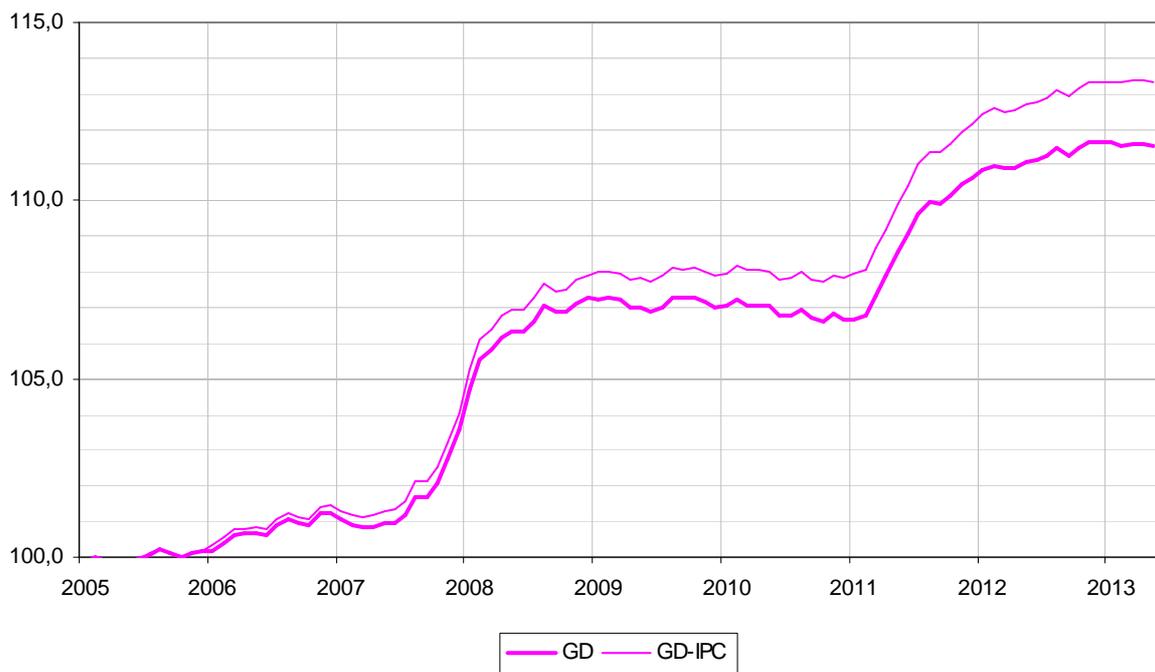
On s'intéresse donc à l'indice des produits de grande consommation, toutes formes de vente confondues, et on compare la série fournie par la méthode IGD-2004 avec celle fournie par la méthode IPC.

Graphique 1.a.1 : Indices (base 100 en moyenne annuelle 2005)
 – Produits de grande consommation
 – Toutes formes de vente



La méthode IGD-2004 produit un indice plus bas que la méthode IPC, lorsqu'on observe les produits de grande consommation, toutes formes de vente confondues. L'écart semble se creuser pendant les périodes de forte inflation. Cette divergence augmente lorsqu'on restreint le champ aux seules observations de la grande distribution, comme le montre le graphique suivant.

Graphique 1.a.2 : Indices (base 100 en moyenne annuelle 2005)
 – Produits de grande consommation
 – Formes de vente : grande distribution

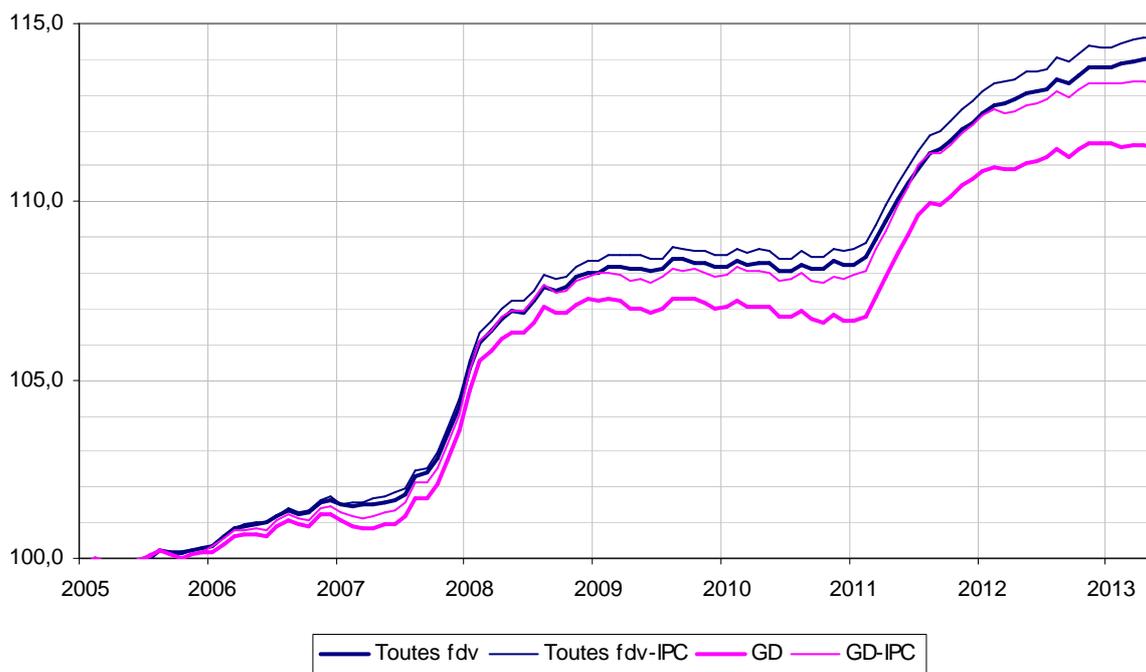


Lorsqu'on superpose les deux graphiques (voir graphique 1.a.3), on s'aperçoit que les deux méthodes mettent en évidence les mêmes sauts historiques, début 2008 et mi-2011, et les mêmes profils d'évolution de prix. Les deux méthodes concluent en tendance à une moindre hausse des prix dans la grande distribution, avec une accélération de l'écart depuis mi-2011.

Toutefois, la méthode IGD-2004 produit, par rapport à la méthode IPC :

- des indices plus bas,
- des écarts grande distribution – toutes formes de vente plus creusés⁵.

Graphique 1.a.3 : Superposition des graphiques 1.a.1 et 1.a.2
Indices (base 100 en moyenne annuelle 2005)
– Produits de grande consommation



Ayant constaté des tendances de long terme divergentes dans les courbes d'indices produites par la méthode IGD-2004 et celles produites par la méthode IPC, il reste à déterminer laquelle des deux méthodes, le cas échéant, est la plus proche de la « vérité ».

1.c.4 DISCUSSION THÉORIQUE SUR LA MÉTHODE IGD-2004

Formellement, on peut reprocher deux choses à la méthode IGD-2004 :

- Elle traite de manière équivalente les variétés homogènes⁶ et hétérogènes. Or la théorie enseigne que la formule de Dutot par exemple, bien adaptée aux variétés homogènes, fournit de mauvais résultats dès lors que les prix à agréger sont hétérogènes et/ou heurtés⁷.
- Elle occulte la dimension géographique de l'inflation. Or, si dans le long terme l'inflation s'équilibre sur le territoire, à un moment donné on peut observer des disparités parfois importantes entre agglomérations.

⁵ Entre 2005 et 2012, l'écart cumulé entre la grande distribution et l'ensemble des formes de vente s'établit à -1,9 points d'indice avec la méthode IGD-2004, mais seulement -0,9 avec la méthode IPC.

⁶ On appelle « homogène » une variété dont la définition délimite une classe de produits très proches en termes de spécification et de qualité de service rendu, par exemple, une boîte de conserves de légumes (disons, des haricots verts), de poids net égoutté compris entre 400 et 600g. Les variétés ne pouvant pas être qualifiées d'homogène selon cette définition sont dites « hétérogènes ». C'est le cas des biens technologiques par exemple, ainsi que des services. Dans le domaine qui nous intéresse ici, à savoir les produits de grande consommation, la plupart des variétés sont homogènes.

⁷ Pour une présentation détaillée des différentes formules de calcul applicables au niveau élémentaire d'agrégation, on pourra se reporter aux chapitres 9 et 20 du *Manuel de l'Indice des prix à la consommation* édité par le FMI (2004).

Pour tenter d'apprécier l'un et l'autre effet, on propose de regarder comment se décompose au niveau des variétés l'écart constaté sur l'indice d'ensemble. Les commentaires qui suivent reposent sur la constatation que la formule utilisée dans la méthode IGD-2004 se rapproche d'une formule de Dutot (pour les variétés homogènes comme hétérogènes).

Tableau 1.a.1 : Comparaison des indices variété obtenus avec les deux méthodes
 – Produits de grande consommation
 – Toutes formes de vente

Type de variété		Écart d'indice moyen (1)	Écart d'indice moyen en valeur absolue	Pourcentage d'écarts positifs
Homogène		0,05	0,11	67 %
	<i>Écart-type</i>	<i>0,01</i>	<i>0,02</i>	2 %
Hétérogène		-0,08	0,36	44 %
	<i>Écart-type</i>	<i>0,01</i>	<i>0,20</i>	1 %

(1) : [méthode IPC]-[méthode IGD-2004]

Pour les variétés homogènes, la méthode IGD-2004 produit un indice plus bas que la méthode IPC, de 0,05 points en moyenne. Ces écarts sont assez peu dispersés, ce qui signifie que cet effet de méthode se constate sur la plupart des variétés homogènes. Or la formule de Dutot est adaptée aux variétés homogènes. L'écart dû à la différence de méthode semble donc bien être due à la seule omission des disparités géographiques.

On notera que les écarts sont plus instables sur les variétés hétérogènes : si, en moyenne, l'écart constaté va dans le sens inverse par rapport aux variétés homogènes, cet écart moyen est entaché d'une dispersion trop importante pour que l'on puisse en tirer une conclusion générale. On notera également que les écarts grimpent plus haut que pour les variétés homogènes, semblant indiquer un caractère relativement décalé de l'usage d'une formule de Dutot pour rendre compte de l'évolution des prix, par rapport à la formule de Jevons classiquement utilisée pour le type de variétés hétérogènes.

Les variétés homogènes étant largement majoritaires dans le sous-champ des produits de grande consommation, c'est leur effet qui l'emporte.

Tableau 1.a.2 : Répartition des types de variété
 – Produits de grande consommation
 – Toutes formes de vente

Type de variété		Nombre de variétés de grande consommation	Poids de ces variétés dans l'IPC d'ensemble
Homogène		257	10,4 %
	<i>– Part dans les variétés de grande consommation</i>	<i>63 %</i>	<i>61 %</i>
Hétérogène		148	6,8 %
	<i>– Part dans les variétés de grande consommation</i>	<i>37 %</i>	<i>39 %</i>

2 Recherche du « meilleur » niveau élémentaire d'agrégation

2.a Pourquoi modifier le niveau élémentaire d'agrégation ?

Les conclusions que l'on peut tirer du § 1.a sont les suivantes :

- Le choix du niveau élémentaire d'agrégation a un impact certain sur les indices produits et également sur les comportements *relatifs* des indices dans la grande distribution et dans les autres formes de vente.
- Dans ce choix du niveau élémentaire d'agrégation, deux effets coexistent : celui lié au choix de la formule de calcul d'une part, et celui lié à la prise en compte (ou non) de la dimension géographique dans les étapes d'agrégation.
- Le comportement des indices de variétés homogènes tend à prouver que c'est l'éviction de la dimension géographique qui tire les indices vers le bas (et non la formule de calcul employée).

Ces différents résultats montrent que le niveau élémentaire retenu dans la méthode IGD est un grain trop épais, qui ne permet pas de capter les disparités régionales en termes d'inflation à un temps t , ceci ayant des conséquences visibles sur l'indice à long terme. On a donc souhaité réintroduire une dimension géographique dans le niveau élémentaire de calcul de l'IGD, tout en gardant à l'esprit que, compte tenu de la taille limitée des échantillons, on ne peut pas conserver l'échelle de calcul élémentaire de la méthode IPC.

Dans la suite, on adopte la méthode de calcul suivante :

- Le niveau d'agrégation élémentaire est réalisé au niveau (variété x groupe d'agglomérations).
- Les étapes d'agrégation suivantes se font par formule de Laspeyres.

On teste plusieurs manières de constituer les groupes d'agglomérations et on apprécie les résultats en fonction de deux paramètres :

- Les séries d'indice calculées selon la nouvelle méthode.
- La robustesse du premier niveau d'agrégation, appréciée à partir de données de fréquences sur le nombre de relevés des cellules élémentaires (variété x groupe d'agglomérations).

2.b Un nouveau découpage du territoire

On aimerait pouvoir regrouper les agglomérations au sein de groupes ayant un sens intelligible d'une part, et présentant une certaine homogénéité en termes de comportement des indices élémentaires d'autre part. Pour cela, on procède d'abord à une analyse des agglomérations en termes de comportement moyen des indices des variétés de produits de grande consommation.

Pour chaque agglomération, on calcule une variable d'intérêt qui consiste en la moyenne, sur la période allant de janvier 2005 à décembre 2012, des écarts pris chaque mois par les indices (variété x agglomération) par rapport aux indices (variété).

On essaye ensuite d'expliquer cette variable d'intérêt par les variables suivantes :

- Position géographique de l'agglomération
- Taille de l'agglomération
- Représentation des différentes formes de vente dans l'échantillon de l'agglomération

Les bases de données de l'IPC contiennent l'identifiant de région et la ZEAT⁸ de l'agglomération, ainsi que la « taille » de l'agglomération en 4 positions : Paris, agglomérations de plus de 100 000 habitants hors Paris, agglomérations entre 20 000 et 100 000 habitants, et agglomérations entre 2 000 et 20 000 habitants.

La variable « taille d'agglomération » semble non significative pour expliquer les différences de comportement des agglomérations en termes d'indices. Seul fait notable : Paris semble avoir un comportement particulier en termes de variance. Ainsi, si l'écart moyen des indices élémentaires

⁸ Zone d'études et d'aménagement du territoire : ce zonage regroupe les régions de France métropolitaine en huit grands ensembles géographiques connexes. Pour une définition complète, consulter le site [insee.fr, rubrique Définitions et méthodes, section Définitions : <http://www.insee.fr/fr/methodes/default.asp?page=definitions/zone-etude-amenagement-territ.htm>](http://www.insee.fr/rubrique/Definitions/et/methodes/section/Definitions)

parisiens s'établit à la faible valeur positive de + 0,04 (valeur non significativement différente de 0), il le fait avec un écart-type deux fois moins important que toutes les autres agglomérations. De cette manière on peut dire qu'il existe un surcroît d'inflation en région parisienne sur les produits de grande consommation, très faible mais assez systématique sur la période considérée.

Cette particularité de comportement conduit à écarter Paris des études suivantes, et à l'isoler dans un groupe qui ne contient que cette agglomération. Cette décision est d'autant plus recevable que le nombre de relevés dans l'agglomération parisienne est très largement suffisant pour assurer une bonne précision des indices élémentaires par forme de vente.

Paris mis à part, la variable ZEAT ne semble pas déterminante pour expliquer les différences de comportement des agglomérations en termes d'indices. Seules les ZEAT « Ouest » et « Méditerranée » passent le test de non égalité des moyennes au seuil de 10% : les indices élémentaires des agglomérations de l'ouest de la France tendent à s'établir en-dessous des indices de variété, en moyenne sur la période 2005-2012, et avec une variance qui est significativement inférieure à celle observée sur les autres agglomérations. À l'inverse, les indices élémentaires des agglomérations du bassin méditerranéen tendent à s'établir au-dessus des indices de variété.

Puisqu'il semble exister un effet géographique, on a souhaité raffiner notre approche de la géographie, et en particulier s'affranchir des découpages administratifs du territoire, qui ne sont pas a priori les plus pertinents pour rendre compte des mouvements économiques.

On a donc récupéré les données de latitude et longitude des 96 agglomérations métropolitaines de l'IPC. Le coefficient de corrélation de Pearson entre l'écart moyen d'indice et la longitude de l'agglomération s'établit à une valeur de + 0,28, que le test de Student associé juge significativement non nulle au seuil de 1%. La latitude, en revanche, semble non significativement corrélée à l'écart moyen d'indice des agglomérations.

Lorsqu'on trace sur une carte les écarts moyens d'indice pour toutes les agglomérations, on remarque un net clivage est-ouest.

À l'est, les zones suivantes présentent des écarts d'indice moyen positifs et significativement supérieurs à la moyenne des autres agglomérations :

- Une zone méditerranéenne coïncidant plus ou moins avec la ZEAT « Méditerranée » (ce qui rejoint donc le résultat précédemment évoqué).
- Une zone incluant les communes de l'est et du sud-est de la région parisienne, à laquelle il faut raccrocher l'agglomération d'Orléans⁹, toutes ces agglomérations présentant la caractéristique d'avoir connu ces dernières années un essor économique lié à la proximité parisienne.
- Une large zone autour de Lyon allant de Clermont-Ferrand⁹ à l'ouest jusqu'à la frontière Suisse à l'est et Grenoble au sud.

À l'ouest, les zones suivantes présentent des écarts d'indice moyen négatifs et significativement inférieurs à la moyenne des autres agglomérations :

- Une zone « ouest » littorale que l'on peut faire descendre jusqu'à l'estuaire de la Gironde (ne coïncidant donc qu'assez partiellement avec la ZEAT « Ouest » déjà évoquée).
- Une zone « centre » regroupant le limousin, le nord de l'Auvergne et le sud de la région Centre.

Toutes les différences de comportement en termes d'indices ne s'expliquent toutefois pas par la simple géographie. Le comportement de certaines petites agglomérations ne coïncide pas avec celui de la moyenne des agglomérations environnantes. Dans certains cas, il a semblé que cette différence pouvait s'expliquer par la composition de l'échantillon des produits dans cette agglomération, en termes de forme de vente : les agglomérations qui ont un plus fort pourcentage de relevés réalisés dans la grande distribution tendent ainsi à avoir des indices élémentaires plus faibles que la moyenne (corrélation significative au seuil de 10%).

Pour apprécier les différents effets et leur possible significativité sur les écarts d'indice moyens calculés, on a procédé par modèles de régression linéaire. La taille d'agglomération regroupée en 4 modalités s'étant révélée peu parlante, on l'a remplacée dans le modèle par des données de population issues du recensement de la population (populations légales de 2010), au niveau de la commune et au niveau de l'aire urbaine.

⁹ Voir par exemple à ce sujet l'analyse en ligne sur le site de la commission nationale du débat public sur la croissance démographique et économique autour des villes d'Orléans et Clermont-Ferrand : <http://www.debatpublic-lgv-pocl.org/informer/base-connaissance-popup.html?rub=83>

Il apparaît d'abord qu'il est impossible d'établir un modèle unique pour l'ensemble des agglomérations. La dépendance des écarts d'indice aux différentes variables indiquées semble dépendre assez fortement de la taille de l'agglomération :

- Pour les petites et moyennes agglomérations (moins de 100 000 habitants), l'effet prédominant est celui de la longitude : plus l'agglomération est située à l'est du territoire, plus ses indices élémentaires tendent à s'établir au-dessus des indices de variété (le coefficient est significativement non nul au seuil de 5%). On peut y ajouter un effet de la part de relevés dans la grande distribution, bien que moins significatif : plus l'agglomération comporte une part importante de relevés dans la grande distribution, plus ses indices élémentaires tendent à s'établir en-dessous des indices de variété.
- Pour les grandes agglomérations au contraire, l'effet prédominant est celui de la taille démographique : plus la population municipale est élevée, plus les indices élémentaires tendent à s'établir au-dessus des indices de variété ; de même, plus l'aire urbaine s'étend en proportion de sa ville-centre (c'est-à-dire, plus le rapport population de l'aire urbaine sur population municipale est élevé), et plus les indices élémentaires tendent à s'élever.

De ces quelques éléments on se gardera toutefois de tirer des conclusions hâtives sur le comportement des différentes régions en termes d'évolutions de prix. En effet :

- Les modèles de régression sont d'assez mauvaise qualité, en ce sens qu'ils n'expliquent qu'une faible part de la variance (les R^2 ajustés valent respectivement 0,08 et 0,13).
- La variable d'intérêt étudiée est une moyenne d'écarts établie sur une longue période. Cette moyenne dissimule une réalité qui va dans le sens d'une très forte volatilité des comportements des agglomérations en termes d'indices. La moitié des agglomérations a changé de signe entre 2005 et 2012. Les très petites agglomérations sont celles qui semblent le plus instable, mais de manière générale, on ne peut tirer aucune conclusion solide des tendances ici observées. La seule chose que prouve cette étude, c'est qu'il n'est pas illégitime de vouloir regrouper les agglomérations en zones de proximité géographique ne coïncidant pas avec les découpages administratifs (région, ZEAT).

À partir de ces éléments, on propose un découpage géographique du territoire en 13 zones.

2.c Les indices par forme de vente obtenus en utilisant le nouveau découpage géographique

Pour servir de point de comparaison dans la suite de cette étude, on a parallèlement procédé à une classification ascendante hiérarchique des agglomérations à partir des seules données d'écarts d'indices moyens.

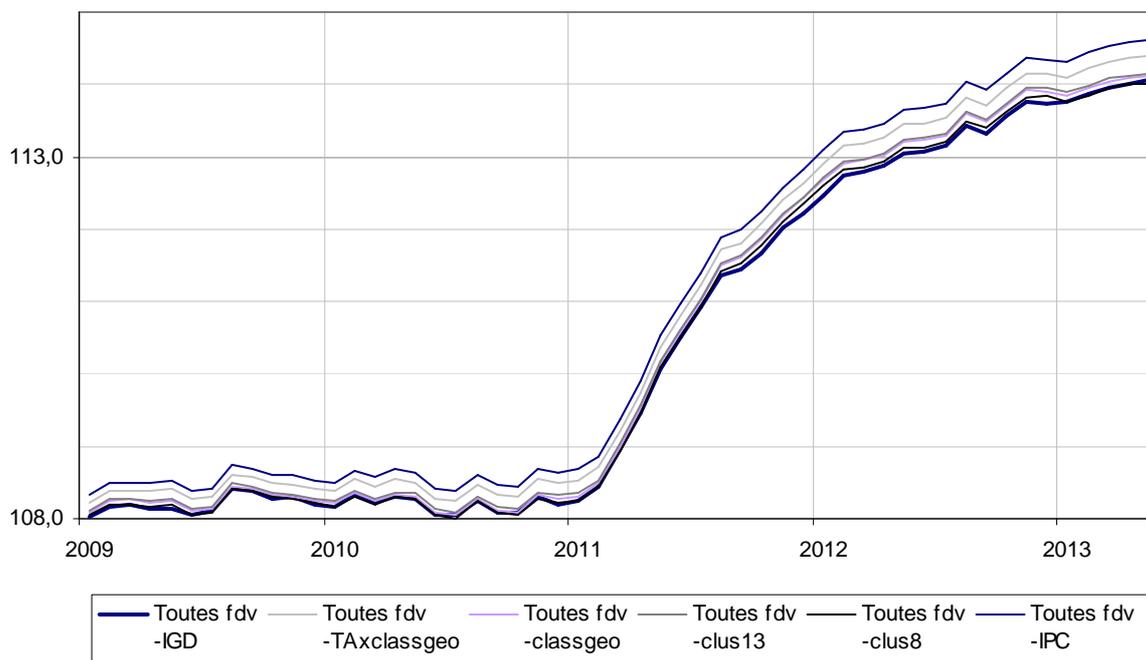
Les classes qui en résultent regroupent des agglomérations homogènes au regard de cette variable d'intérêt dont on a déjà exposé toutes les limites, et n'ont aucun sens autre ; en particulier, les classes obtenues ne recouvrent aucun découpage géographique intelligible, et ne se regroupent pas non plus selon les tailles d'agglomération, par exemple. Par contre, cette classification « statistique » possède sur le découpage en zones géographiques l'avantage d'aboutir à des ensembles beaucoup plus homogènes (toujours en termes d'écarts d'indices). L'écart-type par classe de la variable « écart d'indice » est ainsi de l'ordre de 0,05 points d'indice, alors qu'il est de l'ordre de 0,25 à l'intérieur des zones géographiques.

Cette classification alternative n'a qu'un but statistique, qui est de servir de point de comparaison, permettant d'alimenter la réflexion que l'on aura à propos des résultats auxquels aboutit le découpage géographique proposé dans le paragraphe précédent. À cet effet, on retient deux partitions à partir de la classification ascendante hiérarchique : l'une en 13 classes, l'autre en 8 classes.

Par ailleurs, le passage de 96 agglomérations à 13 groupes d'agglomérations pouvant sembler être une marche importante, on retient également comme candidat le regroupement d'agglomérations issu du croisement entre le nouveau découpage géographique en 13 zones d'une part et la taille d'agglomération d'autre part. Il en résulte une nouvelle partition du territoire en 33 groupes d'agglomérations.

Reprenons le graphique 1.a.1 (toutes formes de vente), superposons les courbes d'indices obtenues avec la nouvelle méthode pour chacun des 4 regroupements d'agglomérations proposés ci-dessus, et zoomons sur la période 2009-2013 pour essayer de mieux percevoir ce qui se passe.

**Graphique 1.b.1 : Indices (base 100 en moyenne annuelle 2005)
 – Produits de grande consommation
 – Toutes formes de vente**



-IGD désigne la méthode IGD-2004

-IPC désigne la méthode IPC, avec niveau élémentaire par (variété x agglomération)

Les autres courbes sont les méthodes alternatives analysées, où le niveau élémentaire d'agrégation est un croisement (variété x groupe d'agglomérations). Le regroupement des agglomérations se fait selon les méthodes présentées précédemment :

- classgeo** : regroupement en 13 zones géographiques étendues
- TAxclassgeo** : croisement de classgeo avec les tailles d'agglomérations, soit 33 classes au final
- clus13** : regroupement en 13 classes issues de la classification ascendante hiérarchique
- clus8** : regroupement en 8 classes issues de la classification ascendante hiérarchique

Entre les courbes de la méthode IGD-2004 et de la méthode IPC, il semble que les courbes des méthodes alternatives viennent se positionner selon un gradient : plus le découpage de premier niveau est fin, plus on se rapproche de la courbe méthode IPC.

Les évolutions sont toutes parallèles, ce qui signifie que quel que soit le découpage retenu au premier niveau d'agrégation, la trajectoire des prix des produits de grande consommation est semblable. La taille du niveau élémentaire d'agrégation a seulement un impact sur le niveau d'indice : sur le long terme, les courbes s'écartent lentement mais régulièrement les unes des autres.

Cet effet ne semble n'être que fonction du nombre de groupes d'agglomérations composant le niveau élémentaire d'agrégation. En particulier, les courbes d'indice obtenues avec le découpage géographique d'une part et la classification statistique en 13 classes d'autre part sont presque confondues. Cela tendrait à indiquer que ce n'est pas le regroupement lui-même qui influe sur le résultat, mais bien la taille des cellules élémentaires (étant entendu qu'on n'a pas regroupé les agglomérations de façon aléatoire). Plus le nombre de classes est élevé, plus la courbe d'ensemble se rapproche de la courbe « méthode IPC ».

Ce résultat invite à retenir le découpage le plus fin possible, sous contrainte que les cellules du niveau élémentaire contiennent suffisamment de relevés pour que la précision de l'indice ne se dégrade pas trop.

Retenons comme candidats possibles les découpages « zones géographiques » (13 positions) et « zones géographiques x TA » (33 positions) et comparons les compositions en nombre de relevés des cellules du niveau élémentaire d'agrégation qui en résultent.

Tableau 1.b.1 : Fréquences des produits dans l'IPC selon la forme de vente
 – Produits de grande consommation
 – Année : 2013

	Toutes formes de vente	Grande distribution	Hors grande distribution
Poids total dans l'IPC	17,2 %	17,1 %	14,1 %
– <i>Poids total perdu dans la forme de vente</i>		0,2 %	3,2 %
Nombre de relevés	41 645	30 016	11 629
– <i>Pourcentage de relevés dans la forme de vente</i>		72,1 %	27,9 %

Les formes de vente hors grande distribution ne représentent que le quart des relevés faits pour les produits de grande consommation. Certaines variétés ne sont pas représentées dans ces formes de vente. On pense donc que c'est l'indice calculé sur les formes de vente hors grande distribution qui va être le plus sensible au problème des cellules élémentaires trop petites.

Tableau 1.b.2 : Fréquences de relevés par cellule élémentaire selon la méthode d'agrégation
 – Produits de grande consommation
 – Formes de vente : hors grande distribution
 – Agglomérations : hors Paris
 – Année : 2013

	Méthode IPC (niveau élémentaire = var-agglos)	Méthode IPC avec regroupmt agglos en 33 positions	Méthode IPC avec regroupmt agglos en 13 positions
Nombre moyen de relevés par cellule	1,6	2,1	3,1
– <i>Ecart-type</i>	0,9	1,9	2,9
Part des cellules ayant moins de 4 relevés	96,8 %	87,0 %	72,8 %
– <i>En pondérations</i>	94,6 %	76,2 %	50,6 %

Lorsqu'on filtre les seules observations des formes de vente hors grande distribution, la quasi totalité des couples (variétés ; agglomérations) comporte moins de 4 relevés. Le risque de dérive de l'indice « autres formes de ventes » est, dans ces conditions, considéré comme trop important pour pouvoir appliquer la méthode IPC.

Regrouper les agglomérations en 13 classes permet de réduire sensiblement cette proportion, même si on reste très loin de l'objectif 0 % : avec le regroupement des agglomérations en 13 zones géographiques, il n'y a « plus que » 73 % de cellules élémentaires avec moins de 4 relevés, ces cellules totalisant la moitié du poids total des produits de grande consommation dans les formes de vente considérées. En d'autres termes, la moitié des cellules élémentaires concourant au calcul de l'indice « autres formes de vente » comprend au moins 4 relevés.

On atteint également, avec cette méthode, un nombre moyen de relevés par cellule élémentaire égal à 3,1 relevés, soit deux fois plus qu'avec la méthode IPC sans regroupement des agglomérations. Ce chiffre se rapproche de la cible théorique qu'on pourrait se fixer en prenant comme référence le nombre moyen de relevés par var-agglos toutes formes de vente confondues, lequel s'élève à 4,5 relevés. Toutefois, l'écart-type important (2,9) relativise ce résultat.

Le choix du regroupement résulte au final d'un compromis entre les deux préoccupations évoquées précédemment : d'une part, on souhaite dans l'absolu se rapprocher le plus possible de ce que donnerait la méthode IPC, lorsqu'on se situe sur le champ « toutes formes de vente » ; d'autre part, la cible étant de calculer des indices par forme de vente, on doit améliorer autant que possible le problème des cellules de trop petite taille.

On propose donc finalement d'opter pour le découpage géographique en 13 zones. Ce découpage présente les avantages suivants :

- Il permet de réduire de moitié la proportion de cellules très petites dans le premier niveau d'agrégation.
- Il obtient des résultats comparables à ceux de la méthode IPC en termes d'évolutions : les courbes sont parallèles même si un décalage s'instaure progressivement.
- Il est intelligible et explicable en termes économiques.

La méthode IGD-2014 est donc la suivante :

Formule 1.a.2 : Méthode de calcul de l'IGD à partir de janvier 2014
– Agrégation par variété

$$I_{m,a}^{FV}(v) = \sum_{\forall zg} \pi_a^v(zg) I_{m,a}^{FV}(v, zg)$$

où

$I_{m,a}^{FV}(v)$ est l'indice de la variété v restreinte à la forme de vente FV , pour le mois m et l'année a ,

$I_{m,a}^{FV}(v, zg)$ est l'indice du couple variété v (restreinte à la forme de vente FV) x zone géographique zg , pour le mois m et l'année a ,

$\pi_a^v(zg)$ est la part de dépense en valeur de la zone géographique zg , pour la variété v et l'année a , indépendante de la forme de vente FV sur laquelle on calcule l'indice,

avec $\forall v, \sum_{\forall zg} \pi_a^v(zg) = 1$ et $\forall v, \pi_a^v(zg) = \sum_{\forall ag \in zg} \pi_a^v(ag)$

Les indices $I_{m,a}^{FV}(v, zg)$ sont les indices de plus bas niveau calculés dans l'IGD. Si la variété est homogène, ils sont obtenus par formule de Dutot ; sinon, par formule de Jevons.

L'indice d'ensemble de la forme de vente FV est obtenue par formule de Laspeyres à partir des pondérations de variétés utilisées dans l'IPC (indépendantes, donc, de la forme de vente FV).