

Utiliser les données de caisses des enseignes de la grande distribution pour le calcul de l'indice des prix à la consommation : quel impact sur l'IPC ?

Les données de caisses des enseignes de la grande distribution sont les données collectées par les enseignes au moment où le consommateur passe à la caisse du magasin pour payer. Sont alors enregistrés les codes-barres des produits achetés, et pour chaque code-barres, le prix et les quantités de produit achetées. Ces données agrégées par point de vente et jour de vente sont ensuite envoyées quotidiennement à l'Insee.

A partir de janvier 2020, les données de caisses des supermarchés et hypermarchés de France métropolitaine sont utilisées pour le calcul de l'indice des prix à la consommation pour le champ des produits alimentaires industriels et des produits d'entretien, d'hygiène et de beauté. Elles remplacent des relevés de prix qui étaient auparavant effectués par des enquêteurs de l'Insee dans les points de vente chaque mois. L'indice des prix à la consommation étant représentatif de l'ensemble de la consommation des ménages (en termes de produits mais aussi de formes de vente), cette collecte par les enquêteurs¹ demeure pour les autres produits vendus en super et hypermarché (par exemple, les produits frais, les biens durables ou l'habillement) et pour les autres formes de vente.

Les données de caisses, un apport majeur pour le calcul de l'IPC

Le recours aux données de caisses pour le calcul de l'IPC est un apport majeur pour son calcul. Outre le prix, il permet de connaître de manière exhaustive les quantités de produits achetées dans cette forme de vente, alors que jusqu'à présent ces informations n'étaient pas disponibles. Cette information est fondamentale pour l'échantillonnage des produits suivis dans l'IPC et l'agrégation des relevés pour produire un indice synthétique. Elle est par ailleurs disponible rapidement, ce qui permet d'identifier précocement les produits nouveaux ou en perte de vitesse et d'adapter en conséquence le panier de produits suivis. La volumétrie des prix suivis permet également d'améliorer grandement la précision de l'indice à ses niveaux les plus détaillés. Dans les données de caisses, les prix sont les prix pratiqués (incluant par exemple les remises individuelles) et non plus les prix affichés collectés par les enquêteurs².

En janvier 2020, seules les statistiques existantes sont produites avec les données de caisses (indice des prix à la consommation, indice des prix dans la grande distribution³), mais les données de caisses devraient à terme permettre de produire de nouvelles statistiques. Parmi celles-ci, on peut citer la production de prix moyens, des comparaisons spatiales de prix plus détaillées et plus fréquentes ainsi que, à plus long terme, des indices régionaux sur un champ restreint de produits.

Des travaux expérimentaux pour garantir l'utilisation de cette nouvelle source de données

Avant d'utiliser les données de caisses, l'Insee a mené des travaux importants tant pour garantir la méthodologie d'exploitation de ces données que l'accès à ces données. Après l'achat d'un premier jeu restreint de données en 2010 permettant de démontrer la pertinence de ces données pour établir des statistiques de prix, l'Insee a reçu quotidiennement, à partir de 2012, à titre expérimental des données de quelques enseignes et a réalisé grâce à elles des études méthodologiques permettant de définir la méthode d'exploitation des données de caisses et l'architecture informatique nécessaire (en technologies big data). Ces travaux se sont enrichis des nombreuses expériences européennes sur le sujet puisque certains pays européens⁴ utilisent déjà les données de caisses pour le calcul de leur IPC.

1 L'IPC utilise également d'autres sources de données que les relevés effectués par des enquêteurs dans des points de ventes physiques comme de la collecte de prix sur internet, le recours à des bases de données administratives...

2 Pour plus de détails sur les apports des données de caisses, voir : Leclair (2019), « Utiliser les données de caisses pour le calcul de l'indice des prix à la consommation », *Le Courrier des statistiques*, n°3

3 Deux nouvelles sous-classes de la nomenclature Coicop seront toutefois suivies grâce aux données de caisses (02.1.3.3 Bière à faible teneur en alcool ou sans alcool et 02.1.3.4 boisson à base de bière) ; certains prix de vente de détail ne seront en revanche plus publiés compte tenu des différences conceptuelles sur les prix moyens pouvant être calculés avec les données de caisses par rapport aux sources de données traditionnelles de l'IPC (prise en compte des quantités en temps réel, prise en compte de manière plus complète des promotions, notamment).

4 Pays-Bas, Norvège, Suisse, Suède, Belgique, Danemark, Islande, Luxembourg, Italie

Par ailleurs, l'accès à ces données de caisses est garanti à l'Insee par la loi. La loi numérique a en effet amendé la loi de 1951 sur l'obligation, la coordination et le secret en matière statistiques et prévoit désormais la possibilité de rendre obligatoire la transmission de certaines données privées, après concertation des acteurs et pour remplacer des enquêtes statistiques obligatoires. Après une concertation avec les enseignes de la grande distribution en juin 2016, le Conseil National de l'Information Statistique (Cnis) a donné un avis favorable à la transmission des données de caisses fin 2016 et un arrêté signé par le ministre le 13 avril 2017 a rendu obligatoire leur transmission pour les commerces de détail en magasins non spécialisés à prédominance alimentaire de plus de 400 m². Depuis janvier 2019, l'ensemble des données de caisses de la grande distribution alimentaire, hors hard discount, à ce stade, sont reçues quotidiennement par l'Insee.

La transmission des données de caisses sur l'ensemble du champ a permis de réaliser une répétition générale du processus tout au long de l'année 2019 et de mesurer, grandeur nature, l'impact de l'utilisation des données de caisses sur la mesure de l'IPC. Les résultats de cette répétition générale sont présentés dans ce dossier. L'impact de l'utilisation des données de caisses sur l'indice d'ensemble en 2019 est relativement faible : -0,08 point au maximum sur l'indice selon le mois, -0,1 point certains mois sur le glissement annuel ou mensuel ; l'inflation mesurée s'avère légèrement plus faible en 2019 avec les données de caisses.

Au-delà de cette répétition générale, l'Insee continuera à suivre tout au long de l'année les données transmises par les enseignes. Des contrôles statistiques seront menés sur les données. Des enquêteurs seront également envoyés dans les super et hypermarchés pour s'assurer, sur un échantillon de contrôle, que les prix transmis via les données de caisses correspondent bien aux prix en magasin.

Le présent dossier revient plus en détail sur la méthodologie utilisée pour traiter les données de caisses et les principales différences avec la collecte traditionnelle (partie 1), puis sur les résultats de la répétition générale (partie 2) et les changements apportés à l'indice de la grande distribution (partie 3).

1 - Une méthodologie d'exploitation des données de caisses conforme aux concepts existants de l'IPC

L'utilisation des données de caisses se fait sans modifier les concepts de l'indice des prix à la consommation, car il s'agit uniquement d'une nouvelle source de données. En particulier, l'IPC avec les données de caisses demeure un indice à panier fixe de type Laspeyres et chaîné annuellement. Les prix d'un panier de biens et services, représentatifs de la consommation des ménages, sont ainsi suivis mois après mois afin de mesurer une évolution de prix à qualité et à structure de la consommation constantes. Ce panier est mis à jour annuellement pour être représentatif de la consommation des ménages. En cas de disparition de produits en cours d'année, ils sont remplacés et un ajustement pour la qualité est réalisé. Les données de caisses permettent cependant d'améliorer grandement le processus en fournissant notamment l'information sur les quantités de produits achetés à un niveau très fin, ce qui nécessite quelques aménagements décrits *infra*. Par ailleurs, le volume très important de données interdit de recourir à un certain nombre de traitements autrefois réalisés manuellement et qui doivent être automatisés.

1.1 Une connaissance fine des achats en super et hypermarché pour définir le panier de l'IPC

Jusqu'à présent, le panier des produits suivis dans l'IPC était défini à partir d'informations de la comptabilité nationale sur le poids des postes dans la consommation. Pour chaque poste, un certain nombre de variétés de produits était retenu sur la base d'informations diverses (sources professionnelles, budget des familles, dire d'experts...). Pour chacune de ces variétés, un certain nombre de relevés de prix étaient effectués, fonction de la volatilité de l'évolution des prix et du poids dans la consommation de la variété. Ces relevés étaient effectués dans des unités urbaines tirées aléatoirement de manière à être représentatives de l'ensemble du territoire. Au sein de ces unités, le choix du point de vente dans lequel le relevé de prix était effectué par l'enquêteur en respectant des quotas par forme de vente. Enfin, au sein d'un point de vente, l'enquêteur retenait un produit parmi tous les produits disponibles pour la variété en privilégiant les produits « bien vendus »⁵.

5 En l'absence d'information précise de l'enquêteur sur la consommation de chaque produit, celui-ci s'appuyait sur la place du produit sur les étagères du rayon et sur l'information que peuvent lui apporter les vendeurs.

Sur le champ des achats en super et hypermarchés, les données de caisses améliorent grandement le processus de choix et de suivi des produits. Précédemment, on ne connaissait pas précisément le détail de la consommation des ménages et un certain nombre de choix n'était contraint qu'au niveau macroéconomique (formes de vente, variété...). Les données de caisses sont une base de sondage et donnent une information objective sur le poids de chaque variété dans le poste (ce qui a amené à revoir l'importance de certaines variétés - les cotons tiges par exemple ne sont plus suivis du fait de leur très faible poids dans la consommation), mais également sur le poids de chaque point de vente et de chaque produit.

1.2 Une meilleure couverture des points de vente et des produits

Outre la possibilité de mieux échantillonner les produits suivis dans le panier et de réduire le biais statistique inhérent, les données de caisses laissent la possibilité d'exploiter leurs données dans leur intégralité, améliorant grandement la précision de l'indice. Compte tenu des technologies informatiques big data, le choix a été fait de retenir un panier quasi-exhaustif. Certains produits ont toutefois dû être exclus : il s'agit des produits éphémères et particuliers ; il s'agit également des variétés dont le poids dans le poste était inférieur à 1 %.

Ce faisant, la couverture géographique est également améliorée puisque tous les points de vente sont retenus, y compris dans les zones rurales ; jusqu'à présent, les prix relevés par l'IPC le sont uniquement dans un échantillon de communes de plus de 2000 habitants. Enfin, les données de caisses comprennent les ventes en drive, qui n'étaient jusqu'à présent pas (ou très peu) suivies dans l'IPC.

1.3 Une nouvelle agrégation des prix

Les données de caisses apportent deux nouveaux types d'information qu'il est nécessaire de prendre en compte dans le calcul d'un indice synthétique : d'une part, les quantités vendues de chaque produit spécifique dans chaque point de vente sont désormais connues ; d'autre part, le nombre de prix observés est désormais bien plus important parce que (i) le nombre de produits est bien plus important (tous les produits dans tous les points de vente contre un seul produit uniquement d'une variété pour quelques points de vente) et que (ii) la fréquence d'observation est bien plus élevée (un prix par jour pour chaque produit si le produit est vendu tous les jours, contre un prix par mois précédemment).

De ce fait au niveau le plus fin, l'agrégation des prix est modifiée (voir tableau 1.1) : faute d'information sur les poids au niveau le plus fin, auparavant, les relevés étaient équipondérés au sein d'une unité urbaine ; ils sont désormais pondérés par leur poids dans la consommation. Par ailleurs, pour un produit donné, le prix pris en compte est le prix unitaire au cours du mois (chiffre d'affaires divisé par les quantités vendues (et le volume du produit)). De ce fait, les promotions, si elles s'accompagnent de quantités vendues plus importantes, ont un effet plus marqué sur les prix que dans la collecte terrain traditionnelle (où leur poids ne dépendait que du nombre de produits en promotion collectés).

Le choix des formules d'agrégation et leur impact sur les résultats de l'IPC sont discutés dans Leclair et alii (2019)⁶.

6 Leclair, M., Léonard, I., Rateau, G., Sillard, P., Varlet, G. et Vernel, P. (2019), « Les données de caisses : avancées méthodologiques et nouveaux enjeux pour le calcul d'un indice des prix à la consommation », *Économie et statistiques*, n°509

Tableau 1.1 : modification des formules d'agrégation pour calculer un indice synthétique des prix à la consommation

Niveau d'agrégation	Collecte terrain	Données de caisses
Prix d'un produit un jour donné dans un point de vente	Il n'existe qu'un seul prix d'un produit dans un point de vente pour une variété donnée	Calculé comme un prix unitaire : chiffre d'affaires du jour divisé par les quantités vendues
Produit donné dans un point de ventes		Calculé comme un prix unitaire : chiffre d'affaires du mois divisé par les quantités vendues
Variété dans un point de vente		Laspeyres géométrique
Variété dans une unité urbaine	Indice de Jevons ou de Dutot	Laspeyres arithmétique
Variété au niveau national	Estimateur d'Horwitz-Thomson utilisant les poids de sondage des unités urbaines	Laspeyres arithmétique
Variété dans un poste	Laspeyres arithmétique	Laspeyres arithmétique

1.4 Un meilleur suivi des prix pratiqués

La nature des prix suivis dans les données de caisses n'est pas la même que dans les relevés effectués par des enquêteurs sur le terrain. L'enquêteur relève le prix affiché, y compris soldes ou promotions si celles-ci s'appliquent à tous et automatiquement à la caisse du magasin. Dans les données de caisses, ce sont en revanche les prix pratiqués qui sont, en général, enregistrés ; les promotions ciblées⁷ sur quelques consommateurs (par exemple les porteurs de carte du magasin) sont, dans la plupart des cas, prises en compte, au prorata de leur poids dans les ventes. En revanche, en l'absence de vente, dans les données de caisses, il n'y a pas de prix : compte tenu du champ actuel d'utilisation des données de caisses, à savoir des produits de très grande consommation, ce point n'est pas problématique.

1.5 Reconnaître un produit identique, une opération automatisée

Le principe d'un indice à panier fixe est que les prix de produits identiques sont suivis mois après mois. Ce faisant, la différence de prix mesurée n'est imputable qu'à l'inflation et à aucun changement de qualité des produits suivis. Ce suivi des produits se fait dans la collecte traditionnelle des prix par l'enquêteur qui s'appuie sur la description du produit qu'il en a faite le mois précédent. Dans les données de caisses, le code-barres du produit est un moyen de s'assurer que le produit est bien identique d'un mois sur l'autre car un code-barres ne peut être utilisé que pour un seul produit. Néanmoins, cet identifiant du produit est un peu trop restrictif, ce qui pourrait mener à sous-estimer l'inflation si on n'adoptait pas les traitements adéquats : le code-barres a en effet tendance à changer lorsque le processus de fabrication est modifié ; parmi ces changements, on peut noter les changements de packaging qui sont souvent accompagnés de hausse de prix mais sans que la nature et la qualité même du produit ne soient modifiées. Ce phénomène de relance commerciale est traité habituellement par les enquêteurs qui décident si le changement apporté est significatif ou non. Dans le cas des données de caisses, une décision similaire est prise et les codes-barres sont liés les uns aux autres sur la base des caractéristiques du produit. Ces caractéristiques sont connues grâce à un référentiel de codes-barres et sont sélectionnées pour chaque variété de produits par un expert du secteur de consommation de manière à définir des produits équivalents (la marque fait en général systématiquement partie des caractéristiques retenues). Ce sont également ces caractéristiques qui permettent de classer les produits en variétés de produits puis dans les postes de la nomenclature utilisée par l'IPC, la Coicop⁸.

7 Les règlements européens, par le passé, excluaient la prise en compte de ce type de promotion (non pas pour des raisons conceptuelles mais du fait de l'impossibilité pratique de les prendre en compte, compte tenu de l'absence d'information sur le pourcentage de ventes en bénéficiant). Le prochain acte d'implémentation prévoit la possibilité de prendre en compte ces promotions pour exploiter les nouvelles possibilités des données de caisses.

8 Classification of Individual Consumption by Purpose

1.6 Remplacer les produits qui disparaissent du panier et ajuster pour la qualité

Dans un panier fixe, lorsqu'un produit disparaît, il est remplacé pour éviter un phénomène d'attrition qui mènerait à un panier de moins en moins représentatif au fil du temps. Le produit remplaçant est choisi le plus semblable au produit disparu mais, si une différence de qualité subsiste, un ajustement qualité est effectué. Dans le cadre de la collecte sur le terrain, ces décisions sont prises par l'enquêteur dans le respect des consignes qui lui sont données. L'ajustement qualité pratiqué repose la plupart du temps sur une méthode de « *bridged overlap* » : la différence de qualité est estimée à partir de la différence de prix entre les deux produits à un même moment. Comme celle-ci n'est pas directement observable (le prix du produit disparu n'est plus observable en t quand on effectue le remplacement, et le prix du produit remplaçant n'a pas été observé en t-1 faute de savoir que le produit initial allait disparaître), on impute le prix du produit remplaçant en t-1 sur la base de l'évolution des prix des produits similaires présents en t et t-1.

Avec les données de caisses, le choix du produit remplaçant se fait aléatoirement parmi les produits de la même variété de produits et dans le même point de vente. Un ajustement qualité est systématiquement pratiqué ; la méthode utilisée est similaire au *bridged overlap* mais le prix passé du produit remplaçant n'a pas besoin d'être imputé, car on peut retrouver a posteriori son prix dans les données de caisses. C'est donc deux prix observés qui sont comparés (deux mois avant la disparition du produit afin de ne pas prendre en compte comme une qualité moindre le fait qu'en général les produits en fin de vie voient leur prix chuter).

2 - Quel impact de l'utilisation des données de caisses sur la mesure de l'inflation en 2019 ?

Afin d'évaluer l'impact sur l'inflation du recours aux données de caisses, un double calcul de l'IPC a été réalisé tout au long de l'année 2019, comme il était réalisé avant 2020 avec des données collectées par les enquêteurs et des données collectées centralement, et comme il sera réalisé à partir de 2020, en utilisant, en sus des autres sources, les données de caisses sur le champ des produits alimentaires industriels et des produits d'entretien, d'hygiène et de beauté pour les super et hypermarchés. Ce double calcul pendant un an permet d'analyser finement les écarts entre les deux indices. L'ensemble des indices habituellement diffusés⁹ et calculés selon les deux méthodes sont disponibles sur le site insee.fr.

2.1 Un panier issu de données de caisses de près de 77 millions de produits

Près de 77 millions de produits ont été retenus dans le panier issu de données de caisses pour l'année 2019, soit plus de 200 000 articles élargis¹⁰ suivis dans plus de 8 000 points de vente de la France métropolitaine. En comparaison, un peu plus de 32 000 relevés effectués par des enquêteurs sur le terrain ont été supprimés de l'échantillon 2019 en raison de leur remplacement par les données de caisses.

Ces articles élargis ont été retenus parce qu'ils étaient vendus au cours de l'année 2018. Ils sont classés en variétés de produits substituables entre eux (même si de qualité différente). Ces variétés ont été définies à l'aide d'un référentiel de codes barres (EAN) décrivant l'ensemble des articles commercialisés en France dans les grandes surfaces alimentaires. Compte tenu notamment de la couverture plus complète des données de caisses, les variétés qui ont été créées pour suivre les produits dans les données de caisses n'ont pas toujours leur pendant dans la collecte terrain, ou leur définition n'est pas forcément équivalente. C'est donc uniquement au niveau supérieur d'agrégation, les postes, que peuvent être effectuées les comparaisons entre les résultats des données de caisses et des relevés terrains.

Au cours de l'année 2019, près de 600 variétés ont ainsi été suivies dans les données de caisses insérées dans 108 postes de la COICOP (graphique 2.1).

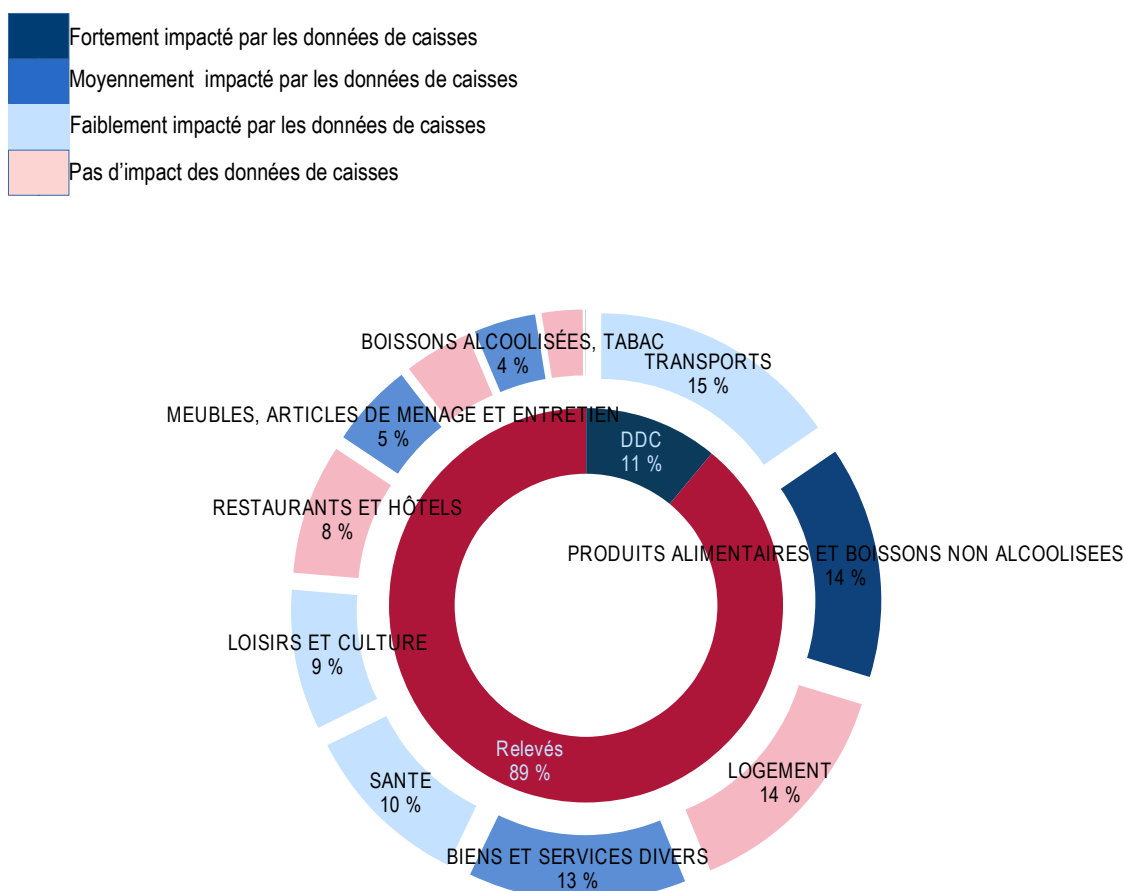
Le champ des données de caisses couvre les hyper et supermarchés de France métropolitaine sur la vente des produits alimentaires industriels, les produits d'entretien, d'hygiène et de beauté. Le panier « données de caisses » ainsi constitué représente 11 % du panier de l'IPC, en termes de dépenses de consommation en 2019¹¹.

⁹ Indice selon la coicop et les regroupements conjoncturels

¹⁰ On appelle un article élargi l'ensemble des codes-barres correspondant aux mêmes caractéristiques.

¹¹ Des travaux complémentaires ont été effectués en 2019 pour réévaluer la part de la consommation sur le champ des données de caisses en confrontant systématiquement les dépenses à un niveau très fin mesurées dans les données de caisses avec les autres sources disponibles (comptes du commerce, comptabilité nationale) ; ces travaux de mise en cohérence des sources amènent à réviser à la baisse cette part des données de caisses : elle ne représente plus ainsi que 9 % du panier de produits suivis dans l'IPC en 2020.

Graphique 2.1 - Répartition de la consommation des ménages (en %) et champ données de caisse



Source : IPC, pondérations 2019, Insee.fr

Note de lecture : en 2019, les produits alimentaires et boissons non alcoolisés représentent 14 % du panier IPC et pour cette fonction, les données de caisses sont fortement utilisées. Les données de caisses représentent 11 % de l'ensemble du panier IPC.

2.2 Des écarts sur l'IPC d'ensemble très faibles

Tout au long de l'année 2019, des indices ont été calculés en utilisant les données de caisses puis ont été agrégés avec les indices traditionnels (indices pour les produits et/ou les formes de ventes non suivis dans les données de caisses) pour calculer un IPC sur le champ complet.

La comparaison de cet indice avec l'indice publié en 2019 montre des écarts relativement faibles¹² (tableau 2.1 et graphique 2.2) : l'écart sur les indices diffusés (arrondis à deux chiffres après la virgule) est au maximum de 0,08 point ; celui sur les glissements mensuels (diffusés arrondis un chiffre après la virgule) est de 0,1 point (effet d'arrondis) en janvier et nul les autres mois ; quant à l'écart sur le glissement annuel, il est nul sauf en juin, juillet, octobre et décembre (écart de 0,1 point). Il est intéressant de noter que l'indice utilisant les données de caisses est systématiquement inférieur à l'IPC construit à partir des relevés terrains. Des écarts de même ordre de grandeur sont observés pour les indices hors tabac (ensemble des ménages, ménages ouvriers et employés, ménages du premier quintile de la population) et les indices hors tabac et hors loyer qui servent pour les indexations.

¹² Au regard des règlements européens, un changement méthodologique est jugé majeur s'il a un impact supérieur à 0,1 point sur l'indice d'ensemble.

Tableau 2.1 – Comparaison de l'IPC publié et l'IPC intégrant les données de caisses
(base 100 = décembre 2018)

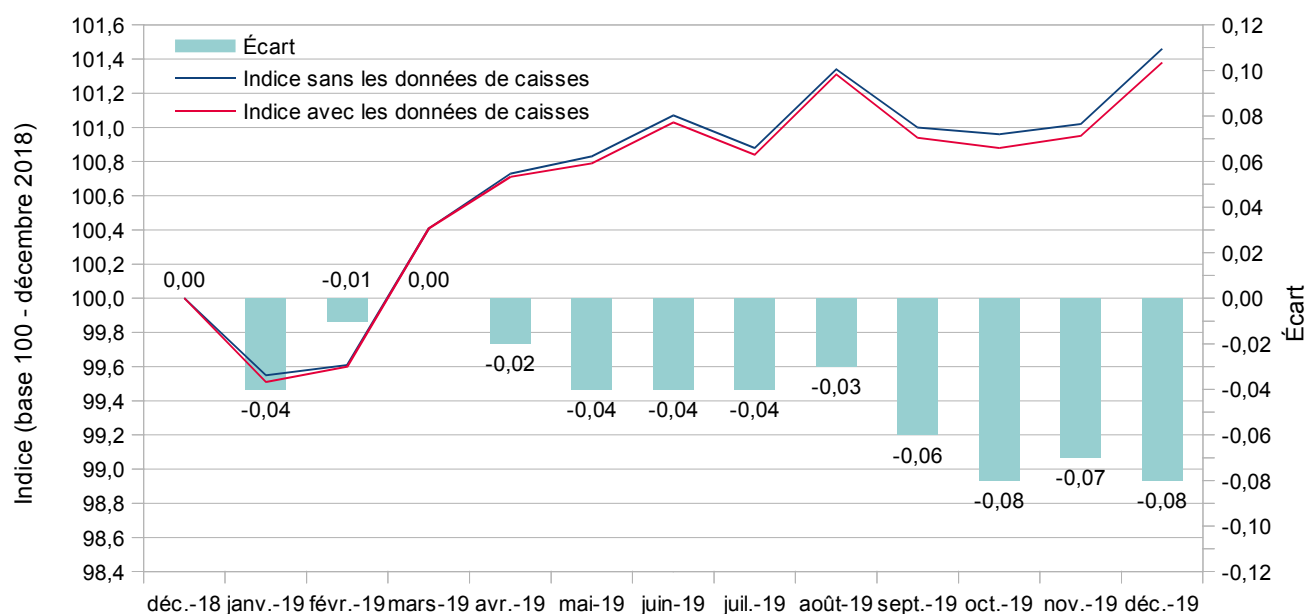
IPC	Indices			Glissement mensuel (en %)			Glissement annuel (en %)		
	Sans les données de caisses	Avec les données de caisses	Écart	Publié	Avec les données de caisses	Écart	Publié	Avec les données de caisses	Écart
Janvier	99.55	99.51	-0.04	-0.4	-0.5	-0.1	1.2	1.2	-
Février	99.61	99.60	-0.01	0	0.1	0.1	1.3	1.3	-
Mars	100.41	100.41	-	0.8	0.8	-	1.1	1.1	-
Avril	100.73	100.71	-0.02	0.3	0.3	-	1.3	1.2	-0.1
Mai	100.83	100.79	-0.04	0.1	0.1	-	0.9	0.9	-
Juin	101.07	101.03	-0.04	0.2	0.2	-	1.2	1.1	-0.1
Juillet	100.88	100.84	-0.04	-0.2	-0.2	-	1.1	1.0	-0.1
Août	101.34	101.31	-0.03	0.5	0.5	-	1.0	1.0	-
Septembre	101.00	100.94	-0.06	-0.3	-0.4	-0.1	0.9	0.8	-0.1
Octobre	100.96	100.88	-0.08	0.0	-0.1	-0.1	0.8	0.7	-0.1
Novembre	101.02	100.95	-0.07	0.1	0.1	-	1.0	1.0	-
Décembre	101.46	101.38	-0.08	0.4	0.4	-	1.5	1.4	-0.1

Note de lecture : indice base 100 en décembre 2018

Champ : France entière ;

Source : IPC, Insee

Graphique 2.2 – Évolution mensuelle des deux indices et écarts par mois.



Note de lecture : indice base 100 en décembre 2018

Champ : France entière ;

Source : IPC, Insee

1.7 2.3 Des écarts plus importants pour les indices détaillés...

L'impact relativement faible de l'utilisation des données de caisses sur les indices d'ensemble est lié en partie au poids modéré du champ données de caisses dans la consommation des ménages (11 %) (Tableau 2.2).

Tableau 2.2 – Comparaison de l'IPC publié et l'IPC intégrant les données de caisses par fonction et principales sous-classes contribuant aux écarts, en décembre 2019 (base 100 = décembre 2018)

INDDEC (Base 100 = décembre 2018)	IPC sans les données de caisses	IPC avec les données de caisses	Ecart	Poids dans la consommation	Poids des données de caisses	Contribution aux écarts
Ensemble	101.47	101.39	-0.08	100%	11%	-
Par fonction et sous-classes						
01 - PRODUITS ALIMENTAIRES ET BOISSONS NON ALCOOLISÉES	102.09	101.69	-0.40	14.3%	44%	-0.06
01.1.2.7 - Viande séchée, salée ou fumée	106.52	105.85	-0.67	1.2%	49%	-0.01
01.1.9.1 - Sauces et condiments	104.24	99.5	-4.74	0.1%	86%	-0.01
02 - BOISSONS ALCOOLISÉES, TABAC	108.94	108.96	0.02	3.8%	39%	0.00
02.1.1.1 - Spiritueux et liqueurs	104.89	105.29	0.40	0.6%	86%	0.00
02.1.3.1 – Bières à basse fermentation	100.55	101.42	0.87	0.1%	83%	0.00
03 - ARTICLE D'HABILLEMENT ET CHAUSSURES	99.75	99.75	-	4.1%	-	-
04 - LOGEMENT, EAU, GAZ, ÉLECTRICITÉ ET AUTRES COMBUSTIBLES	101.18	101.18	-	14.1%	-	-
05 - MEUBLES, ARTICLES DE MENAGE ET ENTRETIEN COURANT DU FOYER	100.20	100.00	-0.20	5.2%	16%	-0.01
05.6.1.1 – Produits de nettoyage et d'entretien	99.87	97.98	-1.89	0.5%	81%	-0.01
06 - SANTE	99.65	99.59	-0.06	10.5%	1%	-0.01
06.1.2.9 – Autres produits médicaux n.c.a. (pansements et compresses)	103.13	101.08	-2.05	0.3%	48%	-0.01
07 – TRANSPORTS	102.46	102.46	-	15.4%	1%	0.00
08 - COMMUNICATIONS	100.71	100.71	-	2.5%	-	-
09 - LOISIRS ET CULTURE	100.51	100.56	0.05	8.6%	5%	0.00
09.5.4.9 – Autres articles de papeterie et de dessin	100.25	101.74	1.49	0.2%	51%	0.00
10 - ENSEIGNEMENT	102.24	102.24	-	0.0%	-	-
11 - RESTAURANTS ET HÔTELS	101.54	101.54	-	8.0%	-	-
12 - BIENS ET SERVICES DIVERS	101.00	100.90	-0.10	13.4%	10%	-0.01
12.1.3.2 - Article d'hygiène corporelle, de bien être et produits de beauté	99.72	99.10	-0.62	1.7%	71%	-0.01

Champ : France métropolitaine

Source: IPC, Insee

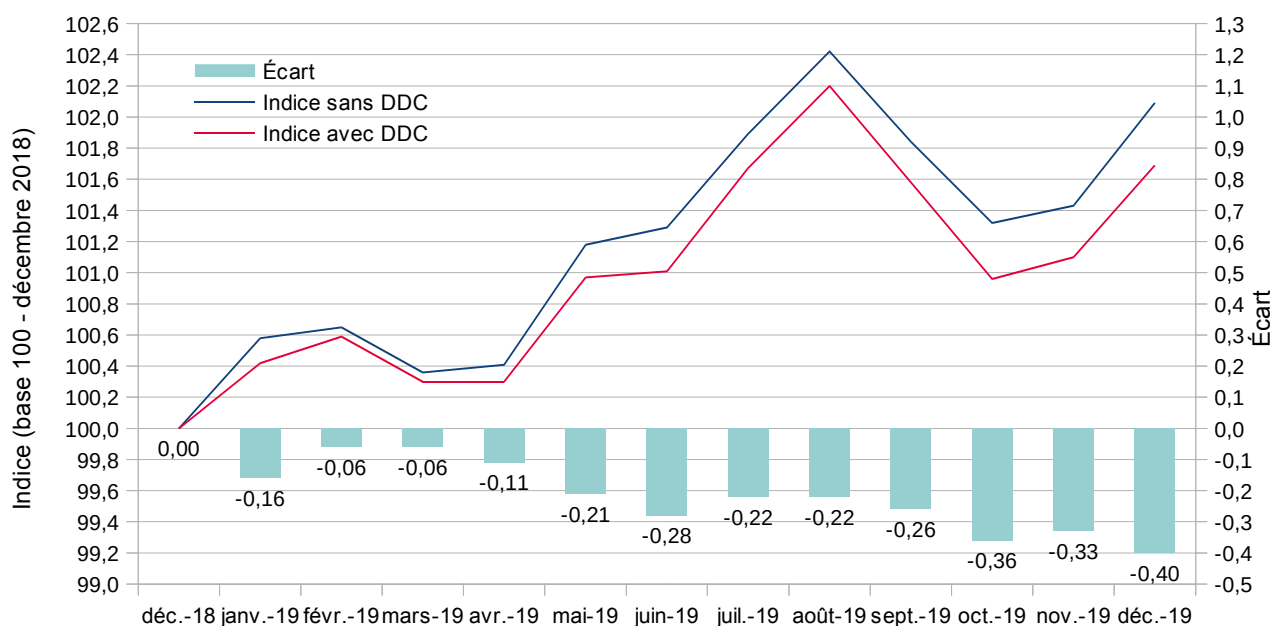
Note de lecture : En décembre 2019, l'indice des produits alimentaires et des boissons non alcoolisés publié sur insee.fr est de 102,9 (exprimé en base 100 en décembre 2018). Il est de 101,69 pour l'indice utilisant les données de caisses, soit un écart de 0,39 point entre les deux indices. Les produits alimentaires et les boissons non alcoolisés représentent 14,3 % de la consommation des ménages. Les données de caisses représentent pour cette fonction 44 %. Cette fonction

L'impact des données de caisses est néanmoins plus important pour les fonctions pour lesquelles le recours aux données de caisses est plus important (tableau 2.2). L'écart sur l'indice de décembre 2019 est ainsi de 0,4 point pour les produits alimentaires et boissons non alcoolisées (fonction 01) pour lesquelles les données de caisses représentent 44 % de l'indice, de 0,2 sur les meubles articles de ménage et entretien courant du foyer (fonction 05) et de 0,1 pour les biens et services divers (fonction 12)¹³.

Pour apprécier l'importance des écarts, rappelons que dans les règlements européens, un écart lié à un changement méthodologique est jugé significatif s'il est supérieur à 0,3 point au niveau de la fonction (et à 0,1 point pour l'indice d'ensemble).

¹³ Les produits d'hygiène et de beauté vendus en hyper et supermarchés sont dans cette fonction.

Graphique 2.3 – Évolution mensuelle des deux indices et écarts par mois pour la division 01 - « Produits alimentaires et boissons non alcoolisées ».



Note de lecture : indice base 100 en décembre 2018

Champ : France métropolitaine

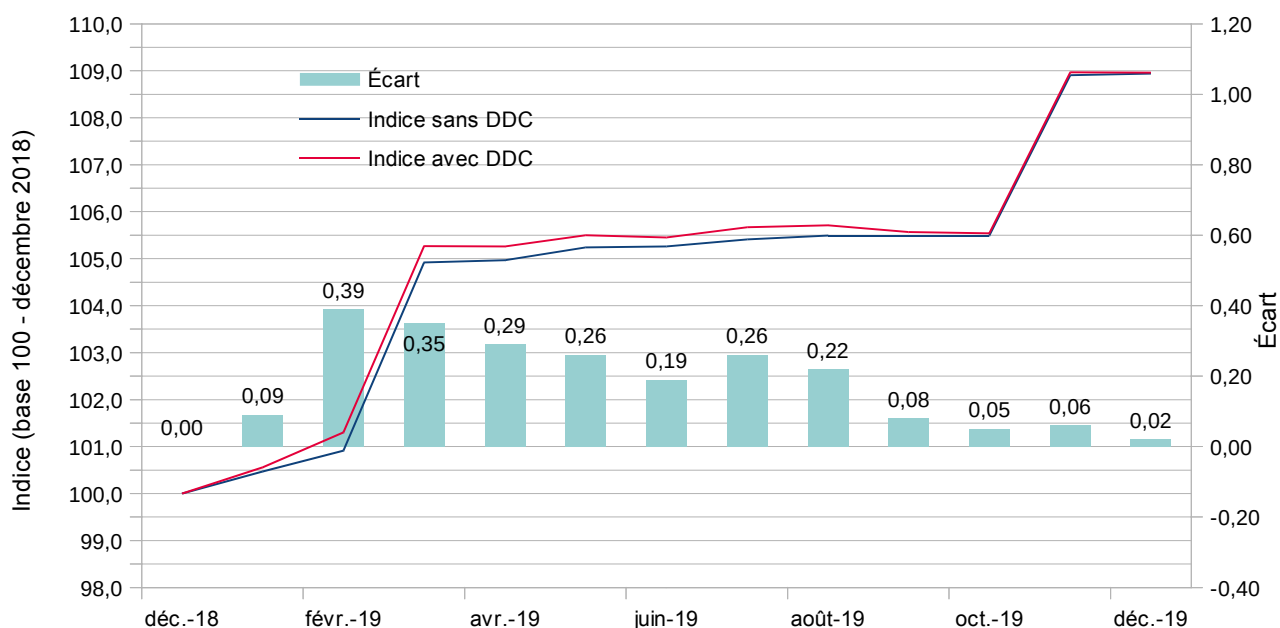
Source: IPC, Insee

Sur le champ des produits alimentaires et boissons non alcoolisées (graphique 2.3), les données de caisses sont utilisées pour tous les produits à l'exception des produits frais (pour lesquels il n'y a pas de codes-barres standardisés), ainsi qu'un certain nombre de viandes et de fromages (notamment quand ils sont vendus à la coupe ou que les poids ne sont pas standards) et du pain, de la pâtisserie ou des viennoiseries produits sur place. Hors des super et hypermarchés, en particulier dans les hard discounts, les supérettes, les marchés et les magasins traditionnels, les relevés sont effectués par des enquêteurs.

Depuis janvier 2019, les écarts entre l'indice des prix des produits alimentaires issu de la collecte terrain et celui intégrant pour partie des indices issus des données de caisses sont compris entre 0,06 et 0,39 (décembre 2019). Les écarts sur le deuxième semestre de l'année 2019 sont plus marqués. Les deux indices donnent des évolutions relativement similaires en termes de glissements mensuels (graphique 2.3).

Sur le champ de l'alimentation, l'indice des prix intégrant les données de caisses est systématiquement en deçà de l'indice des prix issu de la collecte terrain.

Graphique 2.4 – Évolution mensuelle des deux indices et écarts par mois pour la division 02 - « Boissons alcoolisées-tabac ».



Note de lecture : indice base 100 en décembre 2018

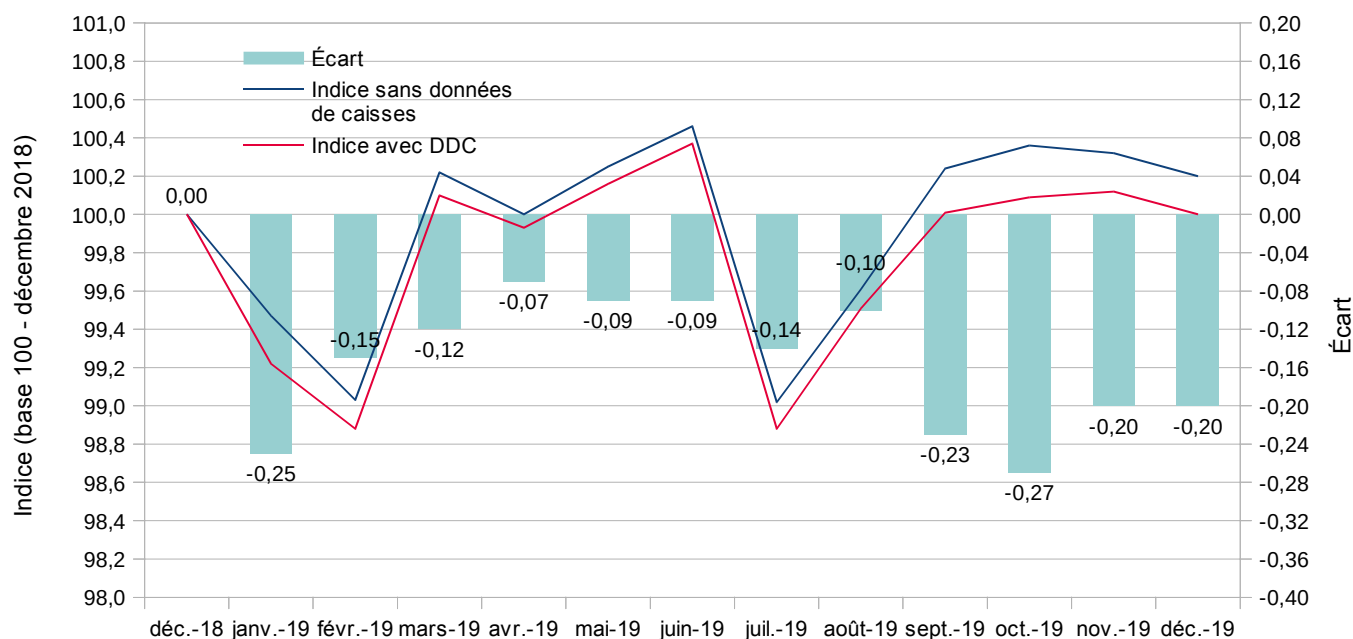
Champ : France métropolitaine

Source : IPC, Insee

Depuis janvier 2019, les écarts entre l'indice des prix des boissons alcoolisées et du tabac (fonction 02) issu de la collecte terrain et celui intégrant pour partie des indices issus des données de caisses sont compris entre 0,02 et 0,39 (février 2019). Contrairement à ce qui est observé sur le champ de l'alimentaire, les écarts entre les deux indices des boissons alcoolisées se sont amenuisés au cours du second semestre, expliquant ainsi l'accroissement des écarts sur l'IPC d'ensemble par non-compensation des écarts négatifs par des écarts positifs.

En effet, sur le champ des boissons alcoolisées, l'indice des prix intégrant les données de caisses est systématiquement au-dessus de l'indice des prix issu de la collecte terrain (graphique 2.4).

Graphique 2.5 – Évolution mensuelle des deux indices et écarts par mois pour la division 05 - « Meubles, articles de ménage et entretien».



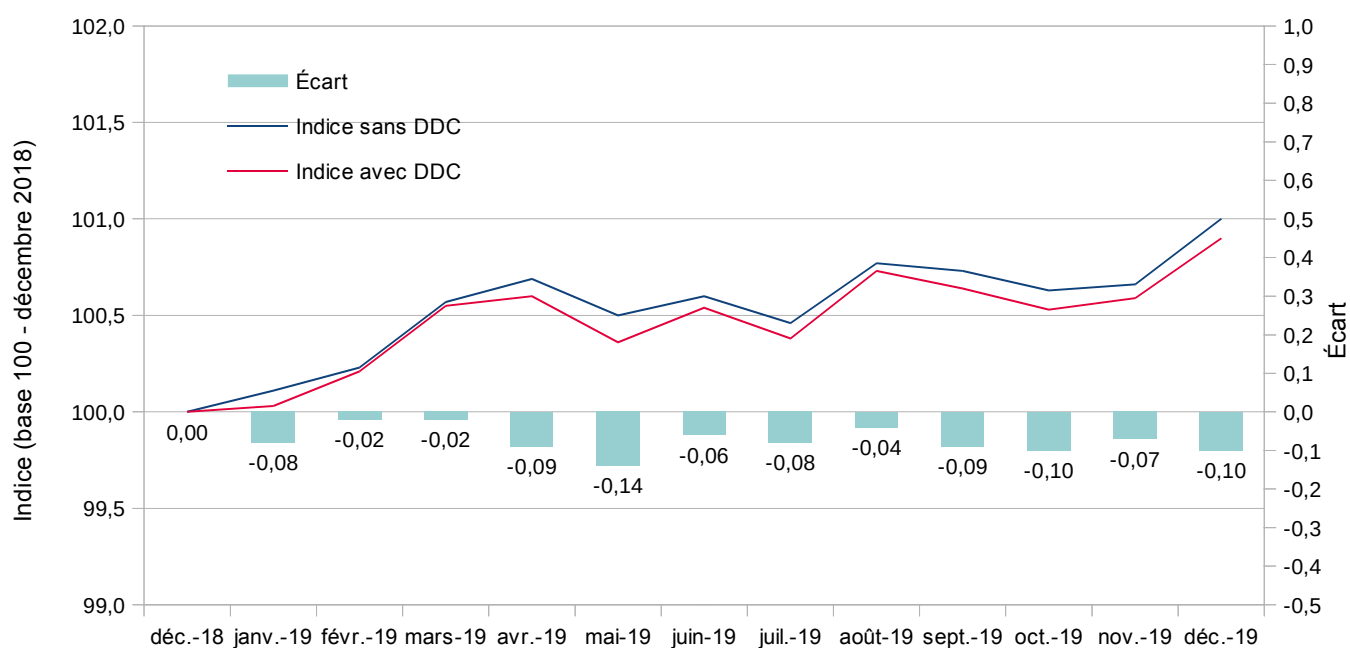
note de lecture : indice base 100 en décembre 2018

Champ : France métropolitaine

Source: IPC, Insee

Depuis janvier 2019, les écarts entre l'indice des prix des meubles, articles de ménages et entretiens (fonction 05, graphique 2.5) issu de la collecte terrain et celui intégrant pour partie des indices issus des données de caisses sont compris entre 0,07 et 0,27 (en octobre 2019).

Graphique 2.6 - Évolution mensuelle des deux indices et écarts par mois pour la division 12 - « Biens et services divers ».



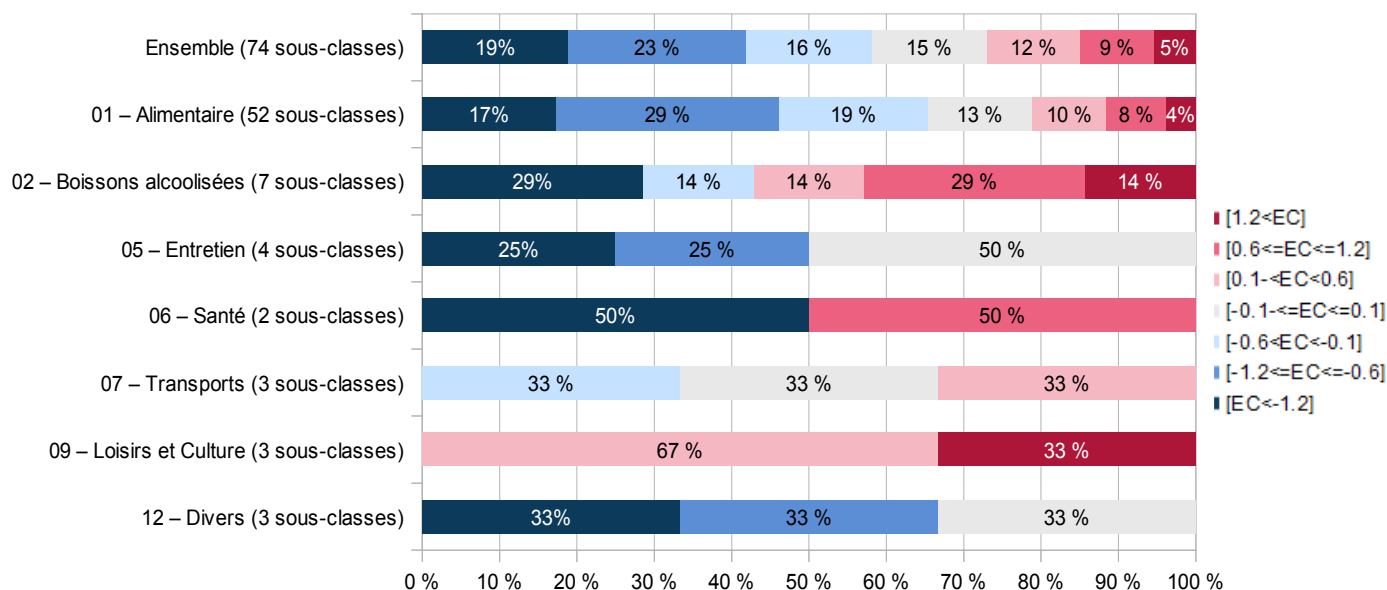
Note de lecture : indice base 100 en décembre 2018

Champ : France métropolitaine

Source : IPC, Insee

Depuis janvier 2019, les écarts entre l'indice des prix des biens et services divers (fonction 12 ; graphique 2.6) issu de la collecte terrain et celui intégrant pour partie des indices issus des données de caisses sont compris entre 0,01 et 0,14 (mai 2019).

Graphique 2.7 - Répartition des écarts (EC) en points entre les indices publiés et les indices intégrant les données de caisses (décembre 2019) au niveau des 74 sous-classes impactées par les données de caisses



Note de lecture : parmi l'ensemble des 74 sous-classes pour lesquelles les données de caisses sont utilisées, 20 % ont des indices utilisant données de caisses qui diffèrent de moins de 0.1 point par rapport aux indices publiés ; 19 % ont des indices inférieurs de 1.2 point en utilisant les données de caisses et 5 % ont des indices supérieurs de 1,2 point avec données de caisses.

Champ : France métropolitaine

Source: IPC, Insee

Au niveau le plus fin de la publication mensuelle (250 sous-classes), 74 sous-classes seront calculées en partie avec des données de caisses : 42 voient leur indice modifié de plus de 0,6 point (ce qui selon le règlement européen est considéré comme un changement significatif à ce niveau de détail), dont 18 de plus de 1,2 point (graphique 2.7).

2.4 ... qui s'expliquent par trois raisons principales

Afin de porter un diagnostic sur la raison de ces écarts, les indices ont été comparés à un niveau très fin.

Trois raisons principales ont été identifiées pour expliquer les différences :

- Les promotions accordées par les enseignes sont en général présentes dans les données de caisses alors que certaines promotions n'étaient pas prises en compte dans les relevés terrain (notamment les promotions discriminantes c'est-à-dire les promotions qui ne s'appliquent pas à l'ensemble des consommateurs de manière systématique ; elles peuvent représenter une part importante des achats). Par ailleurs, le poids des promotions est renforcé par la prise en compte des quantités vendues dans le prix unitaire : elle permet de pondérer les jours avec ou sans promotion par les quantités de produits vendus, les jours de promotion s'accompagnant en général de ventes plus importantes.
- Les indices « données de caisses » sont beaucoup plus précis que les indices terrain du fait du nombre beaucoup plus important de relevés pris en compte. Lorsqu'on travaille sur un échantillon, l'échantillonnage peut conduire à relever malheureusement des prix aux évolutions atypiques.
- Certains postes sont mieux couverts grâce à l'intégration des données de caisses, et plus de segments de consommation sont suivis.

On donne par la suite trois exemples pour illustrer ces phénomènes :

- pour la sous-classe « 05.6.1.1 – Produits de nettoyage et d'entretien », les indices issus des variétés « données de caisses » sont en baisse sur l'ensemble de l'année par rapport à décembre, alors que

les variétés « sœurs »¹⁴ issues de la collecte terrain ont des indices plus stables. L'explication vient du fait qu'il y a eu de nombreuses promotions tout au long de l'année dans les enseignes sur ces types de produits, notamment les produits d'entretien du linge, que ce soit pour les porteurs de cartes de fidélité ou non. A l'inverse, sur le terrain, ces promotions sont moins prises en compte (voir supra).

- sur la sous-classe « 01.1.9.1 - Sauces et condiments », un écart de près de cinq points est observé entre les deux indices au mois de décembre (tableau 2.2). Huit variétés sont suivies par les enquêteurs sur le terrain et treize variétés sont suivies dans les données de caisses. Les huit variétés collectées par les enquêteurs sur le terrain ont toutes une variété "sœur" dans les données de caisses. Une des différences les plus marquées est sur la variété "Moutarde". L'indice publié est de 107,8, celui avec les données de caisses de 99,7. Après analyse, aucune anomalie n'est relevée parmi les 73 relevés collectés sur le terrain dans les hypermarchés, supermarchés et magasins multi-commerce. La valeur de l'indice est affectée par deux articles dont les évolutions de prix par rapport à décembre sont supérieures à 10 %. Ces deux produits représentent 19 % des relevés de la variété. Leur évolution de prix est similaire dans les données de caisses. Cependant, dans le panier « données de caisses », près de 277 000 pots de moutarde sont suivis dans toute la France métropolitaine, ces fortes évolutions de prix constatés sur deux relevés de produits sont donc lissées par rapport aux prix relativement stables des autres articles.
- un écart de 0,4 point est observé entre les indices du mois de décembre pour la sous-classe « 02.1.1.1 - Spiritueux et liqueurs ». Pour cette sous-classe poste, sept variétés sont définies sur le terrain alors que quatorze variétés ont pu être définies avec les données de caisses (âge du whisky, origine du rhum, type de liqueur...). Par exemple, pour le poste « Liqueurs et autres spiritueux non classées ailleurs. », où sur le terrain, les enquêteurs relèvent le prix de produits « bien vendus, bien suivis » comme il leur est demandé ; il s'avère que plus de 50 % des relevés sont des liqueurs de fruits. L'indice de la variété issu de la collecte terrain et l'indice de la variété « Liqueur de fruits » suivie dans les données de caisses sont effectivement relativement proches. Cependant, les évolutions de prix observées sur les autres variétés du poste (liqueur de plantes, liqueur de menthe, crème de fruits, crème de whisky...) ont un impact sur l'indice de prix du poste.

Dans les trois cas, l'utilisation des données de caisses marque un progrès dans la mesure de l'évolution des prix par rapport aux indices précédents, même s'ils ne remettent pas en cause la mesure globale de l'inflation.

¹⁴ Les variétés sœurs désignent des variétés couvrant les mêmes produits entre la collecte « terrain » et les données de caisses.

3 - Un indice des prix de la grande distribution redéfini

L'Indice des prix de la Grande Distribution (IGD) est naturellement plus affecté par l'arrivée des données de caisses puisque son champ est restreint aux produits de grande consommation¹⁵ vendus en super et hypermarché ; il est donc très proche du champ des données de caisses (même si certains relevés terrain sont encore utilisés notamment pour les viandes, le fromage et le pain et viennoiserie, pâtisserie...).

Outre le changement de source de données, l'intégration des données de caisses a entraîné plusieurs changements méthodologiques sur cet indice :

- un changement de champ avec l'ajout des magasins multi-commerces dans le champ de la grande distribution (et dans le même temps, la suppression de ces magasins multi-commerces du champ « hors grande distribution »). L'ajout des magasins multi-commerces n'est pas visible dans l'indice de la grande distribution compte tenu de son poids très faible. Il explique en revanche des modifications dans l'indice hors grande distribution (voir infra)¹⁶, qui n'aurait pas dû être affecté par l'introduction des données de caisses.
- un changement dans la méthode d'agrégation de l'indice, afin d'être plus proche de ce qui est fait désormais dans l'IPC¹⁷ et afin de tenir compte de la structure de consommation par forme de vente¹⁸.

Parmi ces changements, celui ayant le plus d'impact sur la valeur de l'indice de la grande distribution est l'intégration des données de caisses (le changement de source) : les données de caisses représentent en effet 75 % de l'indice de la grande distribution en termes de dépenses de consommation, et 65 % de l'indice de la grande distribution étendue¹⁹ (graphique 3.1). D'après une simulation sur les données 2018, le changement de champ de la grande distribution a un impact de 0,01 à 0,03 point sur l'indice de la grande distribution et jusqu'à 0,15 point sur l'indice hors grande distribution. L'impact de la méthode d'agrégation avait été estimé en 2018 selon les mois entre 0,01 et 0,11 point.

15 Produits alimentaires hors produits frais, boissons, bien d'équipements ménagers non durables et articles et produits pour soins corporels.

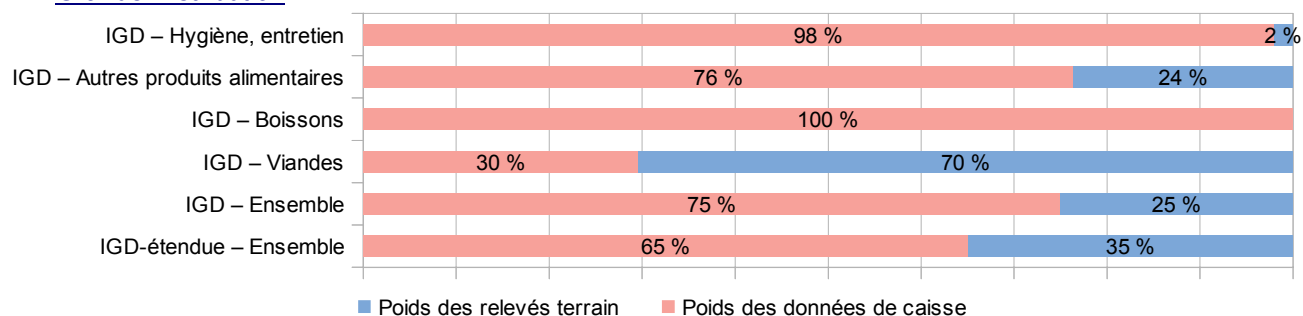
16 L'indice hors grande distribution (« Autres formes de vente ») est publié depuis 2014. Auparavant, on publiait un indice hors grande distribution *étendue*.

17 Compte tenu du faible nombre de relevés par formes de vente, les relevés ne sont pas agrégés par unité urbaine (comme c'était traditionnellement le cas dans l'IPC) mais par taille d'agglomération (alors qu'ils étaient préalablement agrégés pour l'indice des prix de la grande distribution par grande région).

18 Auparavant, faute d'informations détaillées, la structure de consommation par produit pour l'ensemble des formes de vente était appliquée pour chaque forme de vente (alors que les produits consommés sont différents selon les formes de vente). Désormais, les pondérations sont représentatives de chaque forme de vente. Les évolutions des prix de la grande distribution et hors grande distribution peuvent ainsi être différentes pour une raison de structure de la consommation. Il est de toute manière impossible d'imposer une structure identique aux deux indices compte tenu des nombreuses variétés données de caisses suivies pour les hypers et les supers, mais non suivies sur le terrain.

19 L'indice de la grande distribution étendue regroupe les hypers, supers, supérettes et magasins multi-commerce.

Graphique 3.1 – Poids des données de caisses dans les indices publiés dans l'Informations rapides sur la Grande Distribution

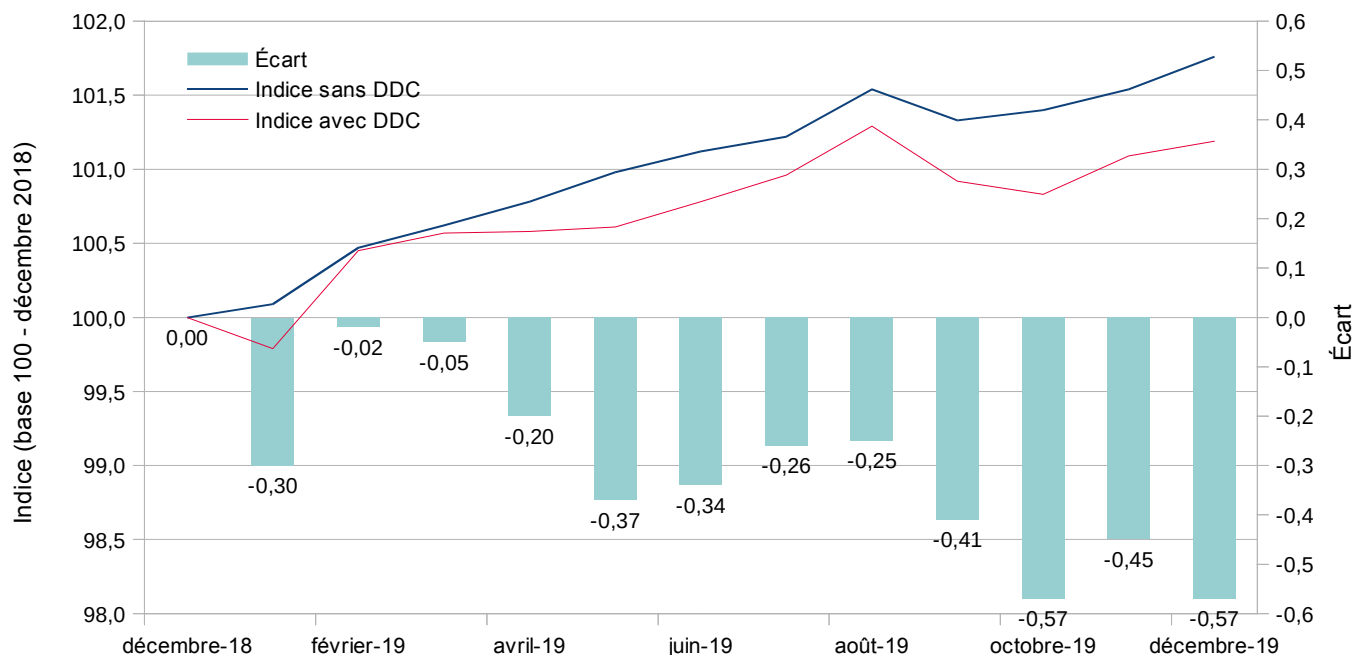


Champ : France métropolitaine
Source : IPC, Insee

Depuis janvier 2019, les écarts observés entre l'indice de la grande distribution publié en 2019 et celui intégrant les données de caisses varient entre 0,01 point (février) et 0,58 point (octobre 2019) (graphique 3.2). Dans le cas de la grande distribution étendue et hors grande distribution, les écarts sont plus faibles (respectivement de 0,01 point à 0,52 point et 0,02 point à 0,25 point, ce dernier étant principalement impacté par le changement de champ). Notons que l'indice de la grande distribution calculé avec les données de caisses est systématiquement inférieur à celui calculé avec des relevés terrain.

Les écarts sur le champ de la grande distribution s'expliquent par les mêmes phénomènes que ceux observés sur l'indice d'ensemble. On note, par exemple, un indice plus faible sur la catégorie hygiène et entretien de la maison : ceci provient de la prise en compte des promotions dans les données de caisses (notamment, celles déjà expliquées précédemment sur la sous-classe « Produits de nettoyage et d'entretien »).

Graphique 3.2 – Comparaison de l'indice de la grande distribution actuellement publié et celui intégrant les données de caisses



Note de lecture : indice, base 100 en décembre 2018
Champ : France métropolitaine
Source : IPC, Insee