

Les données de caisses dans le calcul de l'indice des prix à la consommation (IPC)

FICHE PRESSE – FÉVRIER 2020

Les données de caisses sont les données recueillies par les magasins lorsque les consommateurs passent à la caisse pour payer leurs achats. L'information sur les prix pratiqués chaque jour et sur les quantités vendues pour chaque produit est alors centralisée par certaines enseignes. L'Insee reçoit quotidiennement ces données pour les supermarchés et les hypermarchés. Il s'agit uniquement d'informations agrégées sur les prix et les quantités par produits, points de vente et jours de vente. Aucune information individuelle par consommateur n'est disponible. À partir de janvier 2020 et après plusieurs années d'expérimentations et de tests, l'Insee utilise les données de caisses pour calculer l'indice des prix à la consommation.

À terme, des statistiques nouvelles

En janvier 2020, seules les statistiques produites auparavant par l'Insee sont calculées en utilisant les données de caisses (indice des prix à la consommation, indice des prix dans la grande distribution). Cependant, la richesse des données de caisses permettra, à terme, de produire de nouvelles statistiques de prix sur le champ des produits de grande consommation dans la grande distribution : des prix moyens, des comparaisons spatiales de prix à un niveau géographique plus fin que ce que produisait auparavant l'Insee ou des indices de prix à la consommation régionaux.

Quel est l'intérêt des données de caisses pour le calcul de statistiques de prix ?

Les données de caisses sont des données exhaustives sur les ventes en supermarché et hypermarché et permettent de produire des statistiques **beaucoup plus précises**. Près de **80 millions de produits** sont suivis avec les données de caisses contre 30 000 jusqu'à présent sur le même champ par les enquêteurs de l'Insee. Le prix de ces produits est par ailleurs connu quotidiennement alors que les enquêteurs de l'Insee ne se déplaçaient qu'une fois par mois dans chaque point de vente.

Les données de caisses apportent une information très détaillée sur les quantités consommées par les consommateurs. Ainsi, pour chaque jour, chaque point de vente et chaque code-barres (qui permet d'identifier un produit) la quantité de produit vendue est connue. L'Insee ne disposait pas jusqu'à présent d'une connaissance aussi fine. L'Institut connaissait, par exemple, les dépenses en France en céréales de petit déjeuner mais la quantité achetée dans un supermarché donné pour une marque précise de céréales était inconnue. Grâce à ces informations, la représentativité du panier de produits suivis par l'IPC peut être améliorée et mieux mise à jour en détectant précocement les produits en perte de vitesse ou les produits nouveaux qui apparaissent.

Enfin, **avec les données de caisses, on connaît le prix réel pratiqué par les points de vente**, incluant notamment les promotions. Les enquêteurs ne pouvaient relever que les prix affichés dans les points de vente, incluant les promotions applicables à tous mais pas les promotions personnalisées (pour les porteurs de cartes magasin, par exemple).

Les données de caisses dans le calcul de l'indice des prix à la consommation (IPC)

FICHE PRESSE – FÉVRIER 2020

Quel changement pour la mesure de l'inflation ?

L'utilisation des données de caisses ne change pas les concepts de l'indice des prix à la consommation. Celui-ci reste une mesure de l'évolution pure de prix, à qualité constante des produits et à structure de la consommation donnée. Ainsi une hausse de prix qui accompagne un changement de qualité de produits est neutralisée. De même, si les ménages modifient leurs habitudes de consommation pour privilégier des produits de meilleure qualité (plus de produits bio) par exemple, cela ne sera pas retranscrit dans l'indice des prix à la consommation comme une augmentation de prix.

Quel est l'impact des données de caisses sur l'indice des prix à la consommation ?

Tout au long de l'année 2019, l'Insee a calculé l'indice des prix à la consommation selon deux méthodes : une fois en mobilisant les données traditionnelles de l'IPC (indice diffusé par l'Insee) et une seconde fois en utilisant les données de caisses. Globalement, **la différence observée est très faible** : certains mois, l'inflation mesurée avec les données de caisses est légèrement inférieure (-0,1 point). Cela tient au fait que les concepts de l'IPC n'ont pas été modifiés mais également parce que les données de caisses ne remplacent qu'environ 10 % du panier de l'IPC.

À un niveau plus fin (l'indice des prix à la consommation est produit pour 250 postes chaque mois et 350 postes chaque année), des écarts un peu plus marqués ont pu être observés, en particulier pour les produits alimentaires, les boissons, les produits d'entretien, pour lesquels le recours aux données de caisses est le plus important. Grâce au double calcul de l'indice des prix à la consommation, l'Insee a analysé très finement la raison des écarts poste par poste.

- L'indice avec les données de caisses est beaucoup plus précis ; auparavant, certains produits sélectionnés pouvaient suivre des évolutions de prix atypiques. C'est ce que les statisticiens appellent la « loi des grands nombres » : plus on suit un nombre important de produits et plus on a de chance d'être proche de la moyenne réelle.
- L'indice utilisant les données de caisses couvre mieux l'ensemble de la consommation des ménages ; grâce à la connaissance fine des quantités vendues, de nouveaux segments de consommation ont pu être suivis.
- Enfin, pour certains produits et certains mois, la prise en compte des promotions personnalisées modifie également les évolutions de prix.

L'impact des données de caisses sur l'indice des prix à la consommation est donc faible. Les différences à un niveau fin s'expliquent par un meilleur suivi des prix avec les données de caisses.

Les données de caisses dans le calcul de l'indice des prix à la consommation (IPC)

FICHE PRESSE – FÉVRIER 2020

Comment sont traitées des données si volumineuses ?

Les données de caisses, ce sont près de 80 millions de produits suivis pour le panier de l'IPC mais ce sont aussi 1,7 milliard d'enregistrements reçus chaque mois par l'Insee. En effet, l'information est quotidienne et toutes les données de caisses ne sont pas mobilisées pour l'IPC (les biens durables, l'habillement ou les produits frais, par exemple, continuent d'être suivis par des enquêteurs de l'Insee mais les données de ces produits sont reçues tous les mois).

Pour traiter ce volume de données (« *Big data* »), il faut des **architectures informatiques** particulières. Il faut également être capable d'**automatiser un certain nombre de traitements** autrefois réalisés manuellement. C'est ce qu'a permis de définir toute la phase expérimentale qui a précédé l'utilisation effective des données de caisses.

Les données de caisses sont-elles fiables ?

Oui. Des contrôles statistiques ont lieu tous les jours sur les données reçues et des enquêteurs vont vérifier sur le terrain que les prix mesurés dans les données de caisses correspondent aux prix affichés dans les magasins.

Comme les données de caisses sont des **données privées**, avant de pouvoir être mobilisées pour le calcul de statistiques publiques, un arrêté a rendu **obligatoire** leur transmission à l'Insee par les enseignes de la grande distribution. Cette nouveauté législative, l'obligation de transmission de données privées à des fins de statistiques publiques, a été prévue par la loi de 2016 pour une République numérique.

Auparavant, des enseignes avaient accepté de transmettre leurs données à titre expérimental et l'Insee a pu profiter de ces expérimentations pour effectuer les travaux méthodologiques nécessaires pour les exploiter et s'assurer de la qualité des données.

Les données de caisses remplacent-elles la collecte des prix par les enquêteurs ?

Oui et Non. Ces données **remplacent** les relevés effectués auparavant par les enquêteurs de l'Insee dans les **supermarchés et les hypermarchés** de France métropolitaine pour les produits alimentaires industriels, les produits d'entretien, d'hygiène et de beauté.

L'indice des prix à la consommation couvre toute la consommation des ménages et pas seulement les produits de grande consommation dans la grande distribution. Ainsi, **pour les autres produits** consommés et pour les autres formes de vente, **la collecte de prix habituelle est maintenue** : il peut s'agir de prix collectés par des enquêteurs dans les points de vente ou de prix collectés de manière centralisée (prix sur internet, données administratives, tarifs...).

Est-ce une démarche uniquement française ?

Non. Neuf pays européens, membres de l'Espace économique européen, utilisent déjà les informations issues de données de caisse : les Pays-Bas (depuis 2002), la Norvège (2005), la Suisse (2008), la Suède (2012), la Belgique (2015), le Danemark et l'Islande (2016), l'Italie et le Luxembourg (2018). La quasi-totalité des autres pays européens ont engagé des projets sur le sujet.