



AVIS D'EXAMEN

Enquête expérimentale « Données de caisse »

Service producteur : Institut national de la statistique et des études économiques (Insee), Direction des statistiques démographiques et sociales (DSDS)

Opportunité : avis favorable émis le 18 novembre 2015, par la Commission « Entreprises et stratégies de marché »

Réunion du Comité du label du 2 décembre 2015 (Commission Entreprises).

L'indice des prix à la consommation (IPC) français est produit chaque mois à partir de la collecte de 160 000 prix dans 27 000 points de vente de 106 unités urbaines de métropole et des DOM par 160 enquêteurs de l'Insee. En complément de cette collecte terrain, 40 000 prix sont collectés centralement en bureaux.

L'IPC français est internationalement reconnu pour sa robustesse et sa qualité. Néanmoins, l'essor des références des articles vendus aux consommateurs du fait de la diversification des gammes (marques de distributeurs, produits à bas coût, diététiques, issus de l'agriculture biologique, du commerce équitable, etc.) et le fort développement des ventes promotionnelles à l'initiative des commerçants comme des fabricants rendent souhaitables une augmentation forte de la taille de l'échantillon des produits suivis et une connaissance simultanée des prix et des quantités vendues. Cette extension est confortée par le besoin également exprimé dans le rapport de la commission Quinet « mesurer le pouvoir d'achat des ménages » et par Eurostat.

L'Insee explore toutes les pistes envisageables pour augmenter le volume des prix relevés : *web scraping*, recours aux sources administratives, utilisation des données de caisse. Au niveau européen, plusieurs pays ont d'ores et déjà introduit les données de caisse dans le calcul de l'IPCH : la Norvège (en 1995), les Pays-Bas (2002), la Suisse (2008) et la Suède (2012). La Belgique et le Danemark s'approprient à les utiliser en 2016 ; le Luxembourg et la Pologne en 2017.

L'enquête expérimentale « Données de caisse » a déjà été menée de 2012 à 2015. Quatre enseignes de la grande distribution alimentaire, représentant environ 30 % du marché, y ont participé sur la base du volontariat. La transmission des données était encadrée par des conventions signées avec chaque enseigne. Les données transmises se rapportent aux ventes quotidiennes de chaque article (identifié par son code-barres) dans chacun des points de vente de l'enseigne en France métropolitaine (hors Corse). Elles comprennent, pour chaque article vendu dans un point de vente un jour donné, la quantité d'articles vendue, le prix de vente (et/ou le chiffre d'affaires généré), un court descriptif de l'article (semblable à celui qui figure sur les tickets de caisse) ainsi que l'identifiant de la famille dans laquelle l'enseigne classe l'article.

Un bilan de l'enquête expérimentale a été présenté à la commission *Entreprises et stratégies de marché* du 18 novembre 2015. Un indice expérimental a pu être calculé sur le champ de l'alimentaire, et il montre une bonne proximité avec celui calculé *via* la collecte terrain (réseau d'enquêteurs), sauf bien sûr pour les produits encore largement vendus hors des grandes surfaces alimentaires (exemple de la viande en boucheries traditionnelles). À terme, l'Insee ne s'interdit pas de simplifier les traitements des données – qui peuvent s'avérer très coûteux sur les volumes considérables que

représentent les données de caisse – s'il est prouvé qu'il n'y a pas de différence entre le traitement actuel et un traitement alternatif plus simple et tant que les concepts fondamentaux de l'indice des prix à la consommation sont respectés. En outre, grâce aux données de caisse, des essais de comparaisons spatiales des prix, fondées sur un modèle économétrique dans lequel le niveau de prix est expliqué par un effet fixe spatial, conditionnellement à un effet fixe code-barres, ont été menés. Cette méthode permettrait de réaliser des comparaisons spatiales de prix plus fréquemment, jusqu'alors très coûteuses car fondées sur une enquête terrain monopolisant des enquêteurs.

Pour les années 2016 à 2018, l'objectif principal de l'enquête est de poursuivre le calcul de l'indice expérimental sur le champ des produits industriels alimentaires et de l'étendre au champ des produits d'hygiène-beauté et des produits d'entretien de la maison, de manière à conforter les premiers résultats obtenus et déceler d'éventuels problèmes résiduels. Les indices issus de ce calcul ne seront pas incorporés dans l'IPC diffusé par l'Insee : cette utilisation n'est prévue qu'à partir de 2019 à partir de l'exhaustif des données de caisse. Elle nécessitera une modification de la loi de 1951.

Le second objectif est la poursuite des études méthodologiques : approfondissement des problématiques liées à la saisonnalité, au classement des produits dans la COICOP, à la prise en compte des rabais et promotions, comparaisons spatiales de prix, calculs d'indices régionaux, etc.

Enfin, l'Insee souhaite utiliser les données de caisse collectées pour combler des lacunes sporadiques dans la collecte terrain. Les prix issus des données de caisse s'inséreront dans le calcul de l'indice des prix à la consommation en remplacement de ceux qui n'ont pas pu être collectés par les enquêteurs. L'objectif de ce dispositif est de réduire les imputations de prix dans l'IPC en les remplaçant par le prix réel qui est forcément de meilleure qualité.

L'enquête expérimentale s'étend du 1^{er} janvier 2016 au 31 décembre 2018. Elle concerne les grandes surfaces de la grande distribution sur les produits industriels alimentaires, sur les produits d'hygiène-beauté et sur les produits d'entretien de la maison, toujours sur la base du volontariat.

La collecte est exhaustive et automatique : toute enseigne participant à l'expérimentation fournit l'intégralité de ses données quotidiennes de ventes et de prix (par code-barres et point de vente) pour l'ensemble des produits vendus et pour l'ensemble de ses points de vente situés en France métropolitaine hors Corse. L'Insee collecte les données par voie électronique sécurisée, et par l'intermédiaire d'un tiers de transmission, auprès duquel l'Insee achète des référentiels d'articles et de points de vente.

Le comité de pilotage réunit des représentants des unités de l'Insee concernées par le projet.

Les résultats de l'enquête expérimentale menée de 2012 à 2015 seront par ailleurs présentés fin 2015 aux enseignes volontaires qui transmettent leurs données à l'Insee. De telles présentations seront renouvelées pour les résultats ultérieurs à la convenance des enseignes. L'Insee a œuvré dès le début de l'expérimentation à la mise en place d'un dialogue avec les enseignes.

~~~~~

#### **Le Comité du label émet les recommandations suivantes :**

- Le Comité approuve la mise en œuvre de ce dispositif expérimental qui vise à tester la transmission quotidienne de données émanant d'un ensemble d'enseignes volontaires. Ce dispositif préfigure l'utilisation systématique, à partir de 2019, des données de caisse dans le calcul de l'Indice des Prix à la Consommation (IPC), telle qu'elle pourra être généralisée dans le cadre législatif de la loi « numérique ».

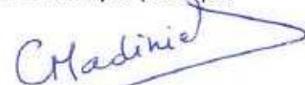
Cette utilisation, à terme, permettra une augmentation globale de la qualité de l'indice des prix grâce à la mobilisation d'un volume plus important d'informations, à la systématisation et à l'automatisation des méthodes, en particulier, celles relatives au remplacement des produits et, éventuellement, à la production d'indices segmentés ;

- Pour toutes ces raisons, le Comité du label considère que cette expérimentation mérite le renouvellement favorable de l'avis d'examen précédemment attribué en 2011 ;
- Le Comité encourage le service à poursuivre les travaux méthodologiques en cours, en particulier, sur les études de saisonnalité. Il recommande au service d'anticiper la transition vers le futur système, notamment en mobilisant les données des autres enseignes dès que la promulgation de la loi « numérique » sera effective, de façon à réaliser un recouvrement des méthodes au minimum sur une année complète : ceci permettra un double calcul de l'Indice des Prix à la Consommation avant le basculement définitif en 2019 sur un nouveau dispositif opérationnel ;
- A plus court terme, le Comité demande de faire figurer le numéro de visa sur les conventions signées avec les enseignes ;
- Le Comité encourage le service à élargir sa communication auprès des enseignes volontaires mais aussi auprès d'autres enseignes, par les canaux qu'il jugera les plus appropriés, afin de les encourager à adhérer progressivement au projet.

**Le Comité du label de la statistique publique attribue un avis d'examen favorable à l'enquête expérimentale « Données de caisse ».**

**Ce label est valide pour les années 2016 à 2018**

La Présidente du Comité du label  
de la statistique publique



Chantal MADINIER