

## Les données de caisse : avancées méthodologiques et nouveaux enjeux pour le calcul d'un indice des prix à la consommation\*

Marie Leclair, Isabelle Léonard, Guillaume Rateau, Patrick Sillard, Gaëtan Varlet et Pierre Vernédal

### Question clé

La statistique publique s'appuie habituellement sur des données d'enquête ou des données administratives. L'accès à des données privées offre de nouvelles possibilités tout en soulevant de nouvelles questions de traitements statistiques. Dans le domaine des statistiques des prix, les données de caisse, données collectées au moment du passage en caisse du consommateur, ont depuis longtemps été exploitées par les enseignes de la grande distribution. L'article montre comment elles peuvent être mobilisées pour le calcul de l'indice des prix à la consommation français.

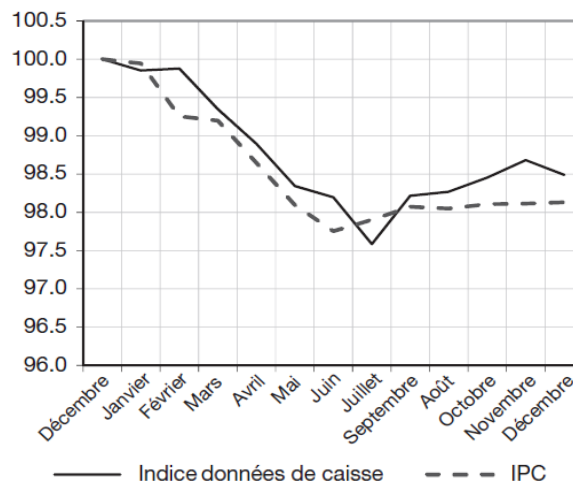
### Méthodologie

L'article s'appuie sur des données de caisse, collectées de manière expérimentale, pour mettre en œuvre les solutions envisagées pour les exploiter (formule d'agrégation des indices, ajustement des effets qualité, etc.). Lorsque différentes méthodes sont possibles, il propose des variantes. Il confronte enfin les résultats de ces exploitations à l'IPC tel qu'actuellement calculé, à partir de relevés effectués par des enquêteurs de l'Insee.

### Principaux résultats

- Les données de caisse permettent de produire des statistiques plus détaillées et plus précises car elles sont exhaustives sur leur champ.
- L'agrégation des relevés de prix pour calculer un indice est modifiée par la disponibilité d'informations nouvelles sur les quantités achetées pour chaque produit ainsi que par le nombre beaucoup plus important de relevés (en continu sur le mois, tous les produits achetés).
- Le remplacement d'un produit du panier peut être automatisé en pratiquant un ajustement qualité ; ce dernier bénéficie de la connaissance rétrospective, dans les données de caisse, des prix passés des produits remplaçants.
- La classification des produits, l'identification des relances commerciales qui étaient faites jusqu'à présent manuellement, peuvent être automatisées en utilisant un référentiel des codes-barres.
- Au total, les indices de prix à la consommation calculés à partir des données de caisse sont proches de ceux calculés actuellement à partir de relevés effectués par des enquêteurs dans les points de ventes.

### Indices des prix à la consommation pour deux postes et indices calculés uniquement sur le champ données de caisse en 2014, base 100 en décembre 2013



### Message

L'utilisation de données de caisse est une opportunité pour le calcul des indices des prix à la consommation. Bien que le volume et la nature des données nécessitent le recours à de nouvelles solutions statistiques (nouveau mode d'agrégation par exemple) et informatiques (recours à des technologies « big data »), l'automatisation d'un certain nombre de traitements et le recours à des référentiels externes enrichissant les données permettent de conserver le cadre conceptuel et méthodologique actuel des indices des prix à la consommation tout en exploitant pleinement les nouveautés des données de caisse (la totalité des données, la connaissance sur les quantités vendues).