

L'utilisation des données de caisses pour le calcul de l'indice des prix à la consommation

Cnis - Commission Entreprises et stratégies de marché - 14 octobre 2020



- **L'indice des prix à la consommation mesure l'évolution des prix à la consommation**
 - d'un panier fixe de produits, pour être à qualité constante
 - mis à jour annuellement, pour rester représentatif de la consommation des ménages
 - en ajustant pour les changements de qualité
- **En s'appuyant tous les mois sur :**
 - 160 000 prix relevés par des enquêteurs dans environ 30 000 points de vente
 - 500 000 prix relevés sur internet (manuellement ou webscraping)
 - Des données administratives
- **Depuis janvier 2020, les données de caisses des super et des hypermarchés (80 millions de produits) ont remplacé :**
 - 30 000 relevés effectués par des enquêteurs en France métropolitaine.
 - sur le champ des produits alimentaires industriels, d'entretien et d'hygiène-beauté

Les grandes dates du projet

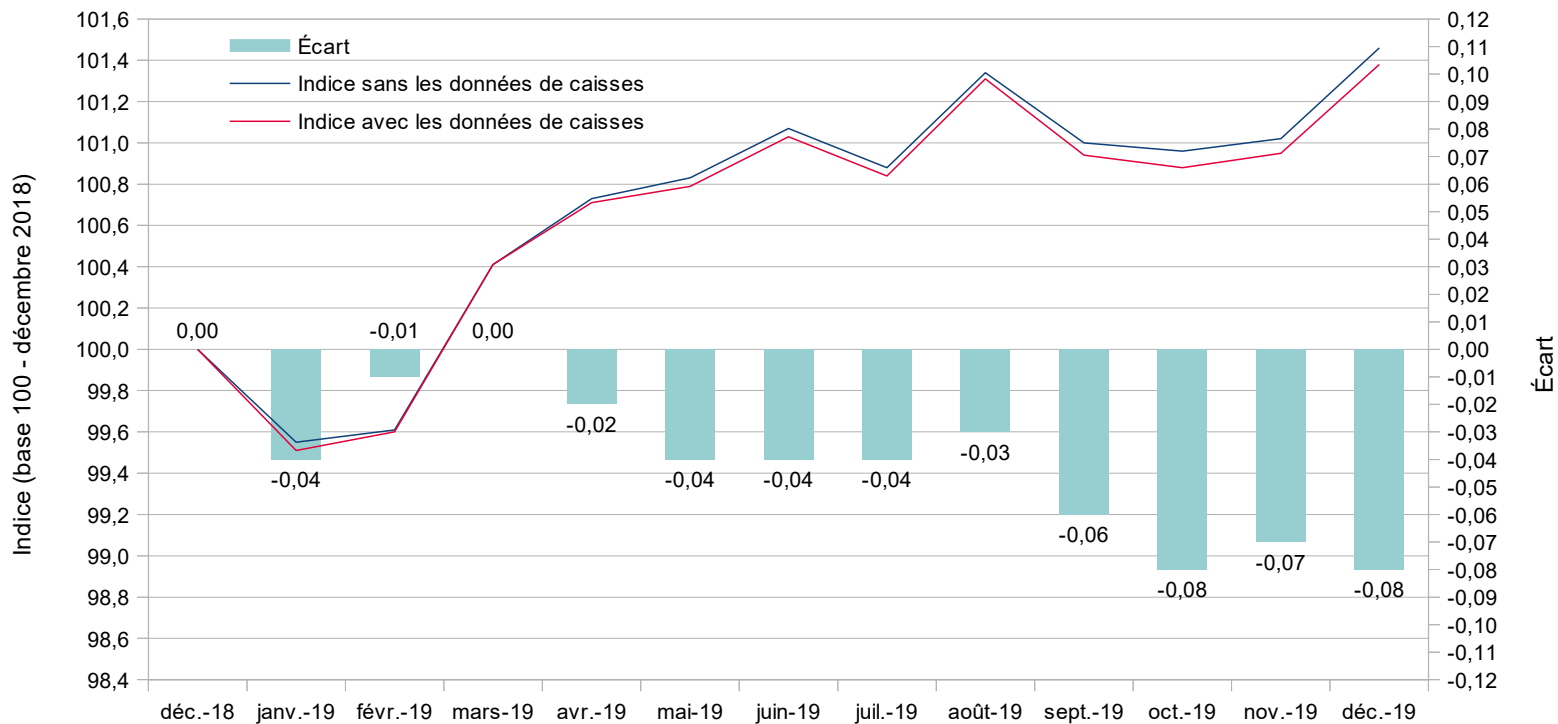


- **Les données de caisses, une information très détaillée :**
 - par point de vente, jour de vente et code-barre
 - sur le chiffre d'affaires et les quantités vendues
 - soit 1,7 milliard d'enregistrements chaque mois.
- **Ces informations permettent de connaître :**
 - l'exhaustivité des ventes en super et hypermarchés
 - les quantités achetées pour chaque produit
 - les prix payés plutôt que les prix affichés

- **Le choix de l'Insee a été de conserver les concepts actuels de l'IPC pour introduire cette nouvelle source de données :**
 - on conserve le principe d'un panier fixe de produits, mis à jour annuellement
 - avec des remplacements de produits en cas d'attrition du panier, et des ajustements qualité
- **Compte tenu du volume de données, cela nécessite de régler quelques difficultés :**
 - avoir une architecture informatique qui permet de tels traitements : recours à une architecture big data (Hadoop).
 - effectuer automatiquement certains traitements faits auparavant de manière manuelle par les enquêteurs (classement des produits, traitement des relances commerciales, des remplacements de produits), grâce notamment au recours à un dictionnaire de codes-barres

Une répétition générale pendant un an pour mesurer l'impact des données de caisses

IPC d'ensemble avec ou sans les données de caisses, (base 100 en décembre 2018)



- **Des écarts inférieurs à 0,1 point sur l'indice d'ensemble mais parfois plus visibles sur les indices détaillés, qui s'expliquent par trois raisons principales :**
 - une précision accrue du fait du nombre de relevés.
 - une meilleure couverture du champ de la consommation
 - une meilleure prise en compte des promotions

- **L'utilisation des données de caisses dans le contexte de la crise sanitaire**
 - Une source qui a permis de calculer des indices pour les produits de grande consommation pendant la période du confinement alors que les enquêteurs ne pouvaient plus se déplacer dans les points de vente physiques
 - Une utilisation accrue des données de caisse :
 - Dans l'IPC, mobilisation des données de caisses des supérettes et pour les biens durables et l'habillement
 - Pour les exercices de nowcasting réalisés par la conjoncture à côté d'autres sources novatrices
- **Par la suite, poursuite des travaux méthodologiques**
 - pour produire de nouvelles statistiques : prix moyens, comparaisons spatiales de prix, IPC régionaux
 - étendre l'utilisation des données de caisses à de nouveaux produits/ à de nouvelles formes de vente / de nouveaux territoires.

Retrouvez-nous sur :

insee.fr



Marie Leclair

Chef de la division des prix à la
consommation

DSDS

marie.leclair@insee.fr

Cliquez pour ajouter un titre