

COMPTE-RENDU DE LA RÉUNION DU : 10/02/2021

Dossier Suivi par :
VOLAT Gwendoline

Tél : 0187696317

Mèl : gwendoline.volat@insee.fr

Montrouge, le 03 mars 2021
N°2021_5087_DG75-F610
Annule et remplace la note N° :

OBJET : Comité de suivi de l'utilisation des données de caisse

RÉDACTEUR : Gwendoline VOLAT et Isabelle LEONARD

TYPE DE COMPTE-RENDU : Pour avis
 Définitif

DESTINATAIRES :

Les participants, Eric Adam (Carrefour), Yannick Bot (Casino), Christel Colin (Insee), Aurélien Daubaire (Insee), Carine Decoudun (Segurel), Eric Dupré (Système U), Sébastien Faivre (Insee), Yoann Foezon (Aldi), Christophe Leray (Les Mousquetaires), S. Maliszewski (Monoprix), Sylvain Moreau (Insee), Laurent Rapoport (Franprix), Anne Russel (Cora), Oriane Verbeke (Match).

Participants : Frédéric Bouttier-Ory (Galec/Leclerc), Edouard de Crozé (Auchan), Romain Dachy (Les mousquetaires), Julie Garbe (Monoprix), Frédéric Jacob (Casino), Adeline Pellerin (Système U), Isabelle Senand (FCD)

Insee : Pascal Chevalier, Alain Jacquot, Pierre Leblanc, Marie Leclair, Isabelle Léonard, Gwendoline Volat

Pascal Chevalier introduit la réunion en remerciant les enseignes pour leur participation à ce comité, sur site ou en distanciel. Il précise l'ordre du jour de la réunion qui débutera par la présentation des travaux utilisant les données de caisses pour les statistiques de prix, puis ceux utilisant les données de caisses pour le calcul d'indices de chiffres d'affaires.

1 - Utilisation des données de caisses pour les statistiques de prix

Gwendoline Volat présente les travaux réalisés utilisant les données de caisses pour les statistiques de prix dans le cadre actuel et les futures utilisations. Elle rappelle le cadre légal de transmission des données (arrêté du 13 avril 2017), puis présente l'utilisation des données de caisses pour l'Indice des prix à la consommation (IPC) publié, effective depuis janvier 2020. En cette année particulière marquée par les confinements successifs et

l'impossibilité pour les enquêteurs de collecter les prix en magasin, les données de caisses ont été un appui précieux pour la production.

Enfin, elle présente les utilisations futures pour les statistiques de prix : enquêtes de parité de pouvoir d'achat, enquêtes de comparaison spatiale des prix, prix moyens, indices de prix régionaux.

Frédéric Bouttier-Ory interroge l'Insee sur l'obtention des données des « hard-discounters », notamment Aldi et Lidl. Pascal Chevalier indique que les discussions se poursuivent avec certains d'entre eux pour établir le flux d'échanges. Pour d'autres, une procédure de contentieux est en cours. Alain Jacquot précise en plus qu'il sera possible d'exploiter les données de caisses des « hard-discounters » et ce sans « catalogues produits » en raison de l'avancée des travaux de machine learning sur les libellés des données de caisses.

Frédéric Bouttier-Ory interroge l'Insee sur l'obtention des données des « pure-players », notamment Amazon. Il précise que cela constitue un vrai « trou dans la raquette » notamment pour les indices de chiffres d'affaires. Il précise que l'année 2020 a permis une forte progression du e-commerce, en citant notamment des données provenant de la fédération de la vente à distance.

Pascal Chevalier précise que le contact est plus difficile, notamment avec Amazon en précisant par ailleurs qu'il ne rentre pas dans le cadre légal actuel.

Frédéric Bouttier-Ory interroge l'Insee sur la prise en compte de la forme de vente « drive » dans les indices issus des données de caisses compte tenu de la possible existence de désynchronisation des prix entre le « physique » et le « drive ». Marie Leclair précise que les ventes de drives sont intégrées et que c'est effectivement une question pour nous de suivre précisément les drives. Cependant, toutes les enseignes ne distinguent pas leurs magasins « drives » de leurs magasins « physiques » ce qui nous conduit à considérer l'ensemble des ventes sur le magasin « physique ». Suite à la remarque de Frédéric Bouttier-Ory sur la transmission des drives Leclerc, Gwendoline Volat confirme leur présence dans les données reçues.

Frédéric Bouttier-Ory demande des précisions sur les enquêtes de parités de pouvoir d'achat. Marie Leclair précise que ce sont des enquêtes conduites sur plusieurs années et ce afin de couvrir l'ensemble de la consommation des ménages, toutes formes de vente confondues. Elle précise également que les comparaisons de prix ne se font pas à l'article précis (même marque par exemple) mais que sont identifiés des articles avec des critères précis pour être comparés (un pack de lessive de même gamme entre les différents pays mais pas forcément de même marque, il faut que ce soit équivalent).

Frédéric Bouttier-Ory demande des précisions sur les prix moyens qui vont être publiés par l'Insee. Marie Leclair précise que l'Insee cherche à répondre à deux objectifs à travers le calcul de prix moyens :

1. Produire des séries historiques (dans la lignée de ce qui était fait par l'Insee avant l'utilisation des données de caisses) de prix moyens sur des produits très homogènes, tel que le prix moyen du lait en €/l, du prix moyen du riz en €/kg...
2. Publier pour un instant t des niveaux de prix moyens pour des produits plus hétérogènes avec un niveau de détail plus fin, par exemple pour le lait, publier le prix moyen du lait entier biologique conditionné en bouteille plastique.

Le niveau de précision des produits publiés respecteront le secret statistique (pas de marque, pas d'identification d'enseignes...)

2 - Utilisation des données de caisses pour les indices de chiffres d'affaires

Isabelle Léonard rappelle que l'Insee a pour projet, depuis quelques années, d'utiliser les données de caisses de la grande distribution alimentaire qui servent au calcul de l'indice des prix à la consommation pour calculer des indices précoces de chiffres d'affaires dans le commerce de détail en remplacement de l'enquête mensuelle d'activité auprès des grandes surfaces alimentaires (Emagsa). L'Insee remercie les enseignes qui ont accepté de participer à l'étude expérimentale. Les résultats de cette étude montrent des évolutions concordantes entre les deux sources et avec la source fiscale sur la TVA. De plus, les données de caisses sont plus précoces que les réponses à l'enquête ; offrent une meilleure couverture ; permettent de produire des indices par grands groupes de produits et par grandes régions. Enfin, elles permettent de réduire la charge statistique pour les entreprises. Les indices publiés respecteront bien entendu le secret statistique et ne permettront pas l'identification des données individuelles. La présente réunion est la première étape du processus légal



pour cette utilisation. L'étude de faisabilité et d'opportunité qui a été transmise aux enseignes invitées et le compte rendu de la présente réunion seront ensuite transmis au Conseil national de l'information statistique et à l'Autorité de la statistique publique pour avis. Si ces deux organismes émettent un avis favorable, un arrêté sera pris par le ministre de l'Économie permettant cette utilisation des données de caisses. Les données et les modes de transmission et de stockage sont identiques à ceux des prix à la consommation. En complément, l'Insee a progressé sur l'utilisation de la source des paiements par carte bancaire qui constitue une source complémentaire à celle des données de caisses avec un champ plus large, en particulier sur les services mais moins de détail sur le commerce de détail alimentaire.

En réponse à une question de Frédéric Bouttier-Ory, l'Insee précise que les grandes régions retenues pour le calcul des indices sont bien les régions administratives et ajoute qu'une attention particulière sera portée aux petites régions dans lesquelles on compte peu de grandes surfaces (elles pourront être regroupées avec une autre région). Des indices par grands groupes de produits ne seront publiés que pour les groupes de produits ayant une part de marché importante.

À la question d'Isabelle Senand sur l'absence des données des carburants dans les données de caisses, l'Insee indique qu'il s'intéresse à d'autres sources en remplacement. Mme Senand fait remarquer que les carburants représentent une part importante dans le chiffre d'affaires des grandes surfaces et que la crise sanitaire a induit des évolutions marquées pour ce produit.

À la question d'Edouard de Crozé sur l'intégration des hard discounters dans les indices de chiffres d'affaires, l'Insee répond que les divisions Prix à la consommation et Indicateurs conjoncturels d'activité se coordonnent sur la démarche engagée pour obtenir ces données.

À la question d'Edouard de Crozé sur la date d'arrêt de l'enquête Emagsa et sur l'information des personnes qui répondent à cette enquête dans les entreprises, l'Insee indique que la fin de l'enquête n'interviendra pas avant la fin de l'année 2022, voire le début 2023 en raison d'un calendrier légal que l'Insee ne maîtrise pas complètement et du besoin de calculer les indices à partir des deux sources sur une période de recouvrement suffisante. La première étape du calendrier légal (présentation à la Commission Entreprises du Cnis) est prévue en mars. L'Insee communiquera régulièrement sur la date de fin de l'enquête.

A la question d'Isabelle Senand sur l'utilisation des données de caisse pour des indices de chiffres d'affaires dans les autres pays européens, l'Insee indique que certains pays du nord de l'Europe ont également lancé des études expérimentales sur l'utilisation des données de caisse ou des données de cartes bancaires pour leurs indices de chiffres d'affaires.

Frédéric Bouttier-Ory fait remarquer que la part de la GSA dans les produits non alimentaires et les services est en baisse au profit des grandes surfaces spécialisées qui sont aussi sur internet. L'Insee partage le constat et précise que l'Insee utilise la source TVA qui contient toutes ces données, mais ne permet pas une publication aussi précoce.

En réponse à une question de Frédéric Bouttier-Ory, l'Insee précise que les personnes de l'Institut autorisées à accéder aux données accèdent aux données élémentaires telles les ventes détaillées d'un magasin précis. Cet accès est indispensable pour la compréhension des évolutions atypiques.

En réponse à une question de Frédéric Bouttier-Ory, l'Insee précise que les données transmises sont hébergées dans un data center propre à l'Insee.

Frédéric Jacob indique qu'il partage les attentes de ses confrères sur l'intégration des hard discounters, du e-commerce et sur la communication d'un calendrier de l'arrêt de l'enquête Emagsa.

