

ECHANTILLON OPTIMAL D'AGGLOMÉRATIONS POUR LA COLLECTE DES PRIX POUR L'IPC FRANÇAIS APRÈS INCORPORATION DES DONNÉES DE CAISSE DE LA GRANDE DISTRIBUTION ALIMENTAIRE

Sébastien Faivre ¹

¹ *Insee, Division des Prix à la Consommation, sebastien.faivre@insee.fr*

Le champ de l'indice des prix à la consommation porte sur l'ensemble des agglomérations de plus de 2000 habitants. Chaque mois, environ 160 000 prix sont relevés par les enquêteurs prix de l'Insee dans plus de 27 000 points de ventes.

La collecte est réalisée en métropole dans un échantillon de 96 agglomérations, stratifié par groupes de région (ZEAT) et par tranches de taille d'unités urbaines (2000 à 20 000 habitants, 20 000 à 100 000 habitants, plus de 100 000 habitants).

L'échantillon d'agglomérations actuel (base 1990) a été tiré sur la base du Recensement de 1990 après un travail d'optimisation mené par Pascal Ardilly et Francis Guglielmetti. Le cadre théorique retenu est celui d'un tirage à deux degrés avec un sondage aléatoire simple à chaque degré de tirage (tirage des agglomérations au sein de chaque strate, puis sélection des produits dont les prix sont relevés au sein de chaque agglomération). Ardilly et Guglielmetti ont alors calculé l'échantillon minimal d'agglomérations à sélectionner par strate (croisement ZEAT et tranche de taille d'agglomération), sous la contrainte de conserver le même niveau de précision que lors de l'échantillon antérieur (base 1970).

L'Insee mène actuellement un projet visant à incorporer dans l'indice des prix les données de prix et de ventes enregistrées aux caisses des points de ventes de la grande distribution alimentaire, en remplacement de la collecte par enquêteurs sur le champ concerné. Ce projet, qui permettrait de disposer de données de prix et de ventes exhaustives sur le champ couvert par les données de caisse tout en rationalisant le dispositif de collecte des prix de l'Insee, conduirait à une réduction globale d'un tiers du nombre de relevés effectués par les enquêteurs.

Une telle variation de la charge de collecte impose de repenser l'organisation de la collecte restante, et de redéfinir l'échantillon d'agglomérations de collecte. Il convient également de vérifier que la collecte restante associée à ce nouvel échantillon d'agglomérations est réalisable sur le terrain, c'est-à-dire qu'elle permet de fournir une charge de travail suffisante à chaque enquêteur.

On présente ici un premier calcul exploratoire "théorique" de l'échantillon d'agglomérations optimal pour la collecte restante, effectué en 2010 dans le cadre de la méthodologie proposée par Ardilly et Guglielmetti. On compare ensuite cet échantillon d'agglomérations "théorique" avec un échantillon d'agglomérations "terrain" issu de considérations sur la faisabilité pratique de la collecte restante. On vérifie alors que le nombre d'agglomérations par tranche de taille dans l'échantillon "terrain" est supérieur à l'échantillon d'agglomérations minimal "théorique", et qu'on conserve donc en concentrant la collecte une précision de l'indice d'ensemble au moins égale à celle de l'indice actuel.