
LA REpondÉRATION DES ENQUÊTES ANNUELLES DE REcENSEMENT POUR UNE DIFFUSION COMPLÉMENTAIRE DU RP

Sébastien HALLÉPÉE, Pierre-Arnaud PENDOLI, Olivia SAUTORY

Insee, Département de la Démographie

sebastien.hallepee@insee.fr
pierre-arnaud.pendoli@insee.fr
olivia.sautory@insee.fr

Mots-clés : sondage, calage, recensement, sources administratives

Résumé

Chaque année depuis 2004, l'Insee réalise une enquête annuelle de recensement (EAR). Chaque EAR est désormais collectée sur la base d'un échantillon, rompant avec le caractère exhaustif des opérations traditionnelles de recensement. Pour les communes de moins de 10 000 habitants, les communes sont toujours recensées exhaustivement une des cinq années du cycle. Pour les grandes communes, le recours à un échantillon pour établir les résultats du recensement est la nouveauté la plus marquante du processus actuel : chaque année, on recense dans chaque commune un échantillon d'adresses représentant 8 % des logements, c'est-à-dire au total 40 % des logements sur un cycle quinquennal.

Les populations légales et les résultats statistiques du recensement de la population (RP) qui sont publiés par l'Insee tous les ans sont issus des données cumulant les EAR d'un cycle complet de RP. Quoique moins précise que le RP en cumul, l'utilisation de l'EAR pour la diffusion de résultats statistiques présenterait toutefois plusieurs avantages puisqu'elle permettrait d'intégrer immédiatement les changements de questionnaire aux résultats publiés. Pour un millésime donné, l'EAR est disponible plus rapidement que les données du RP et ne combine pas plusieurs années d'observation, permettant d'estimer des évolutions plus récentes et d'étudier des phénomènes conjoncturels comme l'emploi.

Cependant le recours à l'EAR demeure actuellement retreint en raison des déséquilibres des groupes de rotation du recensement. Bien que leur création repose sur l'utilisation de plans de sondage équilibrés, les groupes de rotation ne sont pas parfaitement équilibrés au regard des variables auxiliaires d'origine. Ces déséquilibres, qui ont tendance à augmenter au cours du temps, sont à l'origine d'importantes fluctuations des estimations de populations régionales ou départementales d'une année à l'autre.

Pour corriger *a posteriori* ces déséquilibres des groupes de rotation des EAR, une nouvelle méthode de repondération est développée à l'Insee. En amont de l'utilisation de techniques de calage qui permettent de créer les nouvelles pondérations, deux étapes préalables essentielles doivent être considérées.

La première est l'identification des unités de calage, car les plans de sondage de l'EAR diffèrent selon les champs. Pour corriger les déséquilibres associés au plan de sondage initial, il est naturel que les unités de calage coïncident avec les unités d'échantillonnage.

- Pour les communes de moins de 10 000 habitants, dites « petites communes », on cherchera donc à caler un échantillon de communes recensées lors d'une EAR donnée.
- Pour les communes de 10 000 habitants ou plus, dites « grandes communes », on cherchera plutôt à caler les adresses de l'échantillon recensées pour cette EAR.

La seconde est la recherche de sources auxiliaires permettant d'améliorer les estimations d'une EAR. Pour le calage de l'EAR 2018 par exemple, il serait possible de recourir au dernier millésime de RP disponible, i.e. le RP 2015 qui agrège les cinq EAR de 2013 à 2017. Mais cela impliquerait de caler l'échantillon de l'EAR sur des totaux ayant trois ans d'ancienneté. En revanche le Fichier démographique des logements et des individus (Fidéli) rassemble des données exhaustives issues des sources fiscales (cadastre, taxe d'habitation, impôt sur le revenu...) n'ayant qu'une seule année d'ancienneté par rapport à l'EAR. On retrouve en outre dans Fidéli des variables auxiliaires similaires à celles qui ont été utilisées dans les plans de sondage des groupes de rotation, telles que le sexe et l'âge des individus.

La constitution des variables auxiliaires pour chaque unité de l'échantillon n'est toutefois pas immédiate. En effet, si le code de la commune constitue un identifiant commun entre l'échantillon de petites communes et Fidéli, il n'existe pas un tel identifiant entre l'EAR et Fidéli pour l'échantillon d'adresses des grandes communes. Pour être en mesure de créer les variables de calage à partir de Fidéli pour chaque adresse de l'échantillon, un appariement est réalisé avec la base de sondage d'adresses (BSA) des grandes communes dans laquelle l'échantillon annuel est sélectionné. La diversité et la complexité des types d'adresses nécessitent de recourir à un algorithme permettant d'apparier à Fidéli environ 90 % des adresses de la BSA tout en évitant les faux appariements.

Une fois les variables de calage calculées pour toutes les unités de l'échantillon de l'EAR, les paramètres de calage sont choisis de façon à limiter la dispersion des nouvelles pondérations.

À l'issue des traitements, en comparant les estimations de l'EAR avant/après calage, il apparaît que les nouvelles pondérations permettent de réduire les écarts entre l'estimation EAR et le total issu du RP du même millésime. En se plaçant dans une perspective chronologique, on note que les séries d'estimations de populations aux niveaux départemental et régional sont lissées. Les nouvelles pondérations des EAR permettent ainsi de réduire la volatilité des estimations de population que l'on observe initialement d'une année à l'autre. Enfin l'utilisation des nouveaux poids des EAR engendre des gains importants en termes de précision des estimations, en particulier à des niveaux géographiques fins et pour le champ des petites communes.

Bibliographie

[1] Bertrand P., Chauvet G., Christian B., Grosbras J.M., « Les plans de sondage du nouveau recensement », VIII^{èmes} Journées de méthodologie statistique, 16-17 décembre 2002.