

L'ANNÉE 2019 POUR LE PROJET NAUTILE (NOUVELLE APPLICATION UTILISÉE POUR LE TIRAGE DES INDIVIDUS ET DES LOGEMENTS DES ENQUÊTES) ET POUR L'ÉCHANTILLON DE L'ENQUÊTE EMPLOI EN CONTINU

Thomas SAUVAGET¹, Ludovic VINCENT¹, Pierre-Arnaud PENDOLI¹,
Thomas MERLY-ALPA¹ & Sébastien FAIVRE¹

¹ *Insee, 88 Avenue Verdier, 92120 Montrouge {thomas.sauvaget;ludovic.vincent;pierre-arnaud.pendoli;thomas.merly-alpa;sebastien.faivre}@insee.fr*

Résumé. Actuellement et pour encore quelques mois, les enquêtes ménages de l'Insee sont pour la plupart tirées dans le recensement de la population (RP), et réalisées en face-à-face par des enquêteurs. Pour cela, l'Insee a mis en place une méthodologie consistant à sélectionner un échantillon de zones (appelé Échantillon-Maître (EM)), puis, au sein de chaque zone sélectionnée, à échantillonner les logements interrogés.

En 2019, les travaux de renouvellement décennal de l'EM, démarrés en 2016, arrivent à leur terme avec l'entrée en production de l'application Nautile (Nouvelle Application Utilisée pour le Tirage des Individus et des Logements des Enquêtes) d'une part, et du nouvel échantillon de l'Enquête-Emploi en Continu (EEC) d'autre part.

Ces deux renouvellements d'échantillon étaient nécessaires. En effet, côté EM, des études avaient montré un déséquilibre croissant des groupes de rotations du RP au cours du temps, dégradant la qualité de tout EM issu de la méthode actuelle. Par ailleurs, une nouvelle source, Fidéli (Fichier démographique des logements et des individus), a ouvert de nouvelles opportunités. Ce fichier, mis à jour chaque année à partir des fichiers fiscaux, présente les propriétés d'une bonne base de sondage : exhaustivité, unicité, fraîcheur des données. L'utilisation de Fidéli à la place du RP permet de diminuer l'étendue des zones d'enquêtes et d'améliorer leur plan de sondage. Elle rend également possible le tirage d'individus.

Parallèlement, l'échantillon de l'Enquête Emploi en Continu (EEC) arrivait lui aussi à terme, si bien qu'un tirage coordonné avec l'EM a pu être mis en place. Au premier semestre 2019, des travaux complémentaires sur la constitution des grappes de logements à interroger ont permis de finaliser la méthode de mise à jour de l'échantillon qui sera employée au cours des 9 années à venir.

Cette communication présente les derniers développements concernant l'entrée en production de Nautile et du nouvel échantillon de l'EEC au second semestre 2019.

Mots-clés. Échantillonnage, enquêtes ménages, échantillon-maître, enquête emploi en continu, sondage en grappes, équilibrage spatialement réparti.

Abstract. To conduct its household and individual surveys, the French National Institute of Statistics and Economic Studies uses the Master Sample design : a group of areas representing the French territory is drawn, and for each survey a sample is extracted from these areas.

Up to 2019, the Master Sample was built from the annual Census Survey, which is itself based on a

yearly rotation of various subgroups. A new database, Fidéli, assembled from administrative data and tax sources, has recently become available. This database has several good properties : annual release, exhaustivity, and no redundancy. Since the original Master Sample needed to be made up-to-date, the opportunity was taken to design and draw a new one based on Fidéli. This has allowed a reduction in the size of the survey areas, and improvements to the sampling design.

In parallel, the sample of the Labour Force Survey was also in need of being redesigned. A coordinated draw with the Master Sample has then been realised in 2018.

In this presentation we shall describe briefly both samples as well as the recent work done in the first semester of 2019 to allow production to start in the autumn.

Keywords. Sampling, Household survey, Labour Force Survey, Master sample, cluster sampling, spatial sampling.

1. Contexte : Nautile et Fidéli

Grâce à une plus grande disponibilité et une meilleure qualité des sources administratives issues des services fiscaux, l'Institut National de la Statistique et des Études Économiques (Insee) a développé une alternative pour ses bases de sondage : le Fichier Démographique sur les Logements et les Individus (Fidéli), qui regroupe les informations fiscales, corrigées des doublons et des règles de gestion administrative, voir Lollivier (2015). Il présente en définitive toutes les caractéristiques d'une bonne base de sondage, en particulier l'exhaustivité.

L'utilisation de Fidéli à la place du recensement de la population, traditionnellement utilisé comme base de sondage à l'Insee (voir à ce sujet Faivre (2009)), permet de diminuer l'étendue des zones d'enquêtes et d'améliorer leur plan de sondage. Cependant, ce changement entraîne la disparition d'informations, la modification de concepts et l'apparition de nouvelles variables, notamment pour le repérage des enquêtes qui pourra se baser sur une géolocalisation plus précise.

Des travaux préliminaires de Favre-Martinoz et Merly-Alpa (2016) ont permis de montrer que la méthode de sondage spatialement équilibré apportait des bénéfices importants pour le tirage d'un Échantillon-Maître. En effet, comme la méthode permet de limiter la sélection d'unités proches géographiquement et donc partageant des caractéristiques socio-économiques proches, elle améliore la précision de variables, en particulier celles n'ayant pas pu être incluses dans les variables d'équilibrage ; elle favorise également une moindre détérioration de la précision des variables dans le temps.

Pour réduire le nombre de variables d'équilibrage, la méthode mise en œuvre, et présentée dans Guillo (2018), a été de synthétiser l'information contenue dans l'ensemble des variables par des méthodes d'analyse de données. L'application d'une analyse en composantes principales au jeu de données au niveau unités primaires a ainsi permis de faire émerger une quinzaine d'axes expliquant 99 % de l'inertie du nuage de données. L'équilibrage sur ces axes permet alors de gagner en précision dans l'estimation des variables du nuage ainsi que de celles qui y sont corrélées, tout en respectant bien mieux la contrainte de taille fixe.

Le renouvellement de l'Échantillon-Maître dans le cadre du projet de Nouvelle Application Utilisée pour le Tirage des Individus et des Logements des Enquêtes (Nautile) fait notamment appel à un tirage spatialement équilibré s'inscrit cependant dans un cadre plus large : celui d'une coordination avec l'échantillon de l'Enquête Emploi en Continu (EEC).

2. Coordination avec l'échantillon de l'EEC

L'enquête Emploi est une enquête visant à observer à la fois de manière structurelle et conjoncturelle la situation des personnes sur le marché du travail. Il s'agit du pendant français de la "Labour Force Survey" (LFS). En France, il s'agit de la seule source fournissant une mesure des concepts d'activité, de chômage, d'emploi et d'inactivité tels qu'ils sont définis par le Bureau International du Travail. Cette enquête est aréolaire : son échantillon est un ensemble de zones géographiquement compactes appelées grappes, regroupées en secteurs. Une grappe est un ensemble d'une vingtaine de logements, et chaque secteur contient six ou sept grappes. Chaque trimestre, une grappe de chaque secteur est interrogée exhaustivement dans un délai de deux semaines ; au bout de six interrogations de la même grappe, elle est remplacée par la suivante au sein du même secteur. Ce mode d'interrogation, mis en place par Loonis (2009) pour une durée de 9 ans, est reconduit pour le futur échantillon EEC.

Contrairement à ce qui a été présenté plus haut, l'échantillon de l'enquête Emploi n'a pas vocation à être représentatif sur un large panorama de sujets : cette enquête se concentre sur les thématiques liées au marché du travail, et il est nécessaire d'être précis uniquement sur ces points, éventuellement avec des déclinaisons selon des domaines de diffusion. Pour plus de détails sur les travaux réalisés par l'Insee à ce thème, voir Costa (2018).

La coordination des tirages de l'EM et l'échantillon de l'EEC a été étudiée afin de limiter les déplacements des enquêteurs. Les unités primaires de l'Échantillon-Maître sont tirées par échantillonnage spatialement équilibré ; l'échantillon de l'enquête Emploi est un ensemble de grappes compactes d'une vingtaine de logements également sélectionnées par sondage spatialement équilibré sur des variables proxy de la situation vis-à-vis du marché du travail. Enfin, leur coordination se fait par l'introduction d'unités de coordination sélectionnées par sondage indirect via les unités primaires, les grappes emploi étant sélectionnées dans les unités de coordination dans une dernière phase. Pour plus de détails, consulter Paliot (2018).

3. Entrée en production au second semestre 2019

L'application Nautile sera livrée prochainement. Un premier tirage aura lieu fin 2019 : il s'agit de celui de l'enquête Technologie de l'Information et de la Communication (TIC).

Concernant l'échantillon de l'EEC, une série de travaux sur la constitution des grappes de logements à interroger en terme de Résidences Principales, Résidences Secondaires et Logements neufs, ont permis de finaliser une méthode de mise à jour des grappes qui sera employée tout au long des 9 années de durée de vie de l'échantillon.

Bibliographie

Lollivier S. (2015), Le répertoire statistique des logements, Commission Territoires du CNIS, https://www.cnis.fr/wp-content/uploads/2017/09/DC_2015_1re_reunion_COM_Territoires_RLS.pdf

Faivre, S et Christine, M. (2009), Le projet OCTOPUSSE de nouvel Échantillon-Maître de l'Insee, Actes des Journées de Méthodologie Statistique de 2009, Insee.

Favre-Martinoz, C. et Merly-Alpa, T. (2016), Utilisation des Méthodes d'Échantillonnage Spatialement Équilibré pour le Tirage des Unités Primaires des Enquêtes Ménages de l'Insee, 9eme Colloque Francophone sur les Sondages, Gatineau.

Guillo, C. et Merly-Alpa, T. (2018), Un nouvel Échantillon-Maître pour 2020 et pour Nautile, Actes des Journées de Méthodologie Statistique de 2018, Insee.

Loonis, V. (2009), La construction du nouvel échantillon de l'Enquête Emploi en Continu à partir des fichiers de la Taxe d'Habitation, Actes des Journées de Méthodologie Statistique de 2009, Insee.

Costa, L., Merly-Alpa, T. et Chevalier, M. (2018), Le renouvellement de l'échantillon Emploi : améliorations et évolutions, Actes des Journées de Méthodologie Statistique de 2018, Insee.

Paliot, N., Chevalier, M. et Deroyon, T. (2018), Coordination spatiale d'échantillons : application à l'EEC et l'Échantillon-Maître, Actes des Journées de Méthodologie Statistique de 2018, Insee.