

Évaluation de la correspondance du lieu de résidence du recensement et des données fiscales à l'aide d'un couplage de données

Julien Bérard-Chagnon et Georgina House¹

Résumé

Les données fiscales sont de plus en plus exploitées pour la mesure et l'analyse de la population et de ses caractéristiques. Un des enjeux soulevés par l'utilisation croissante de ce type de données concerne la définition du concept du lieu de résidence. Si le recensement se base sur le concept du lieu habituel de résidence, les données fiscales nous renseignent sur l'adresse postale des déclarants. Au moyen d'un couplage d'enregistrements entre le recensement, l'Enquête nationale auprès des ménages et les données fiscales du fichier T1 sur la famille, cette étude examine le niveau de correspondance du lieu de résidence de ces deux sources et les caractéristiques qui lui sont associées.

Mots Clés : Recensement; données fiscales; lieu de résidence; adresse; correspondance.

1. Introduction

Les données fiscales sont de plus en plus exploitées pour la mesure et l'analyse de la population et de ses caractéristiques. Ces données permettent non seulement de soutenir différents processus statistiques comme des enquêtes mais sont aussi utilisées directement dans l'élaboration de plusieurs produits de données et d'analyses. Si cette approche revêt plusieurs avantages, elle pose également un certain nombre de défis. L'un des principaux enjeux consiste en la définition du concept du lieu de résidence. En effet, bien que plusieurs programmes tels que le recensement et les estimations démographiques se basent sur le concept du lieu habituel de résidence, les données fiscales nous renseignent sur l'adresse postale des déclarants. Or, le lieu de résidence constitue un élément fondamental de l'étude des populations. En effet, la majorité des indicateurs statistiques utilisés pour éclairer les grands enjeux socio-économiques s'appuient sur la capacité des sources de données à placer les gens au « bon endroit » (National Research Council, 2006). Les différences conceptuelles entre le recensement et les données fiscales à ce chapitre sont donc susceptibles d'influer considérablement sur la comparabilité des deux sources et donc, des produits qui en découlent.

Cette étude examine le niveau de correspondance du lieu de résidence de ces deux sources et les caractéristiques qui lui sont associées au moyen d'un couplage d'enregistrements entre le recensement, l'Enquête nationale auprès des ménages (ENM) et les données fiscales du fichier T1 sur la famille (T1FF). La section suivante traite du concept de lieu de résidence et des différences entre le recensement et le T1FF à ce sujet. Par la suite, la section 3 introduit le couplage de données qui a été utilisé pour cette analyse. La section 4 présente le niveau de correspondance du lieu de résidence entre les deux banques de données pour différents échelons géographiques. Enfin, la cinquième section examine le niveau de correspondance selon différentes caractéristiques individuelles.

¹Julien Bérard-Chagnon, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6 (julien.berard-chagnon@canada.ca); Georgina House, Statistique Canada, 100, promenade Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6 (georgina.house@canada.ca).

2. Concept de lieu de résidence

S'il peut paraître simple à première vue, le lieu de résidence est un concept très complexe. Bien que la majorité de la population soit capable d'établir son lieu de résidence avec un haut niveau de certitude, cette information est plus difficile à cerner pour plusieurs personnes. Les enfants en garde partagée de même que certains étudiants et travailleurs alternent régulièrement entre plus d'un lieu de résidence tandis que les personnes sans-abris n'ont, par définition, pas de lieu de résidence. De plus, certains individus peuvent percevoir leur lieu de résidence non pas comme l'endroit où ils passent le plus de temps mais comme l'endroit avec lequel ils maintiennent les liens économiques ou sociaux les plus forts. Cette situation peut par exemple s'appliquer aux jeunes adultes qui quittent le domicile familial pour la première fois. Les difficultés à établir le lieu de résidence constituent d'ailleurs l'une des principales sources d'erreurs de couverture au recensement (Statistique Canada, 2015 : 30).

2.1 Différences dans le concept de lieu de résidence entre les données fiscales et le recensement

Il n'existe pas de définition canonique du lieu de résidence. Ce concept varie d'une source à l'autre en fonction de l'utilisation principale des données. Ces objectifs parfois divergents font en sorte qu'un même individu peut avoir un lieu de résidence différent d'une source à l'autre.

D'une part, les recensements canadiens utilisent un concept de résidence *de jure*. En conséquence, l'adresse réfère au lieu habituel de résidence, qui est défini comme le logement au Canada où une personne habite la plupart du temps². Cette approche est nécessaire à une bonne planification des services communautaires comme les écoles et le transport public, l'allocation de fonds aux différents paliers gouvernementaux et le redécoupage des circonscriptions électorales. Le formulaire censitaire comprend des règles pour les cas où le lieu habituel de résidence est plus difficile à établir. À titre d'exemple, les étudiants qui retournent vivre chez leurs parents durant les vacances doivent être dénombrés au domicile familial même s'ils passent une partie importante de l'année, voire la majeure partie, ailleurs.

D'autre part, les données fiscales nous renseignent sur l'adresse postale des déclarants³. Cette adresse est requise afin que l'Agence du revenu du Canada (ARC) puisse communiquer avec le déclarant. En ce sens, elle ne vise pas à déterminer où le déclarant habite la plupart du temps mais plutôt un endroit où le rejoindre efficacement. Un déclarant fiscal peut donc très bien inscrire une adresse postale qui n'est pas son lieu habituel de résidence. Ceci pourrait par exemple être le cas d'une personne très âgée dont la déclaration fiscale est remplie par un membre de la famille ou par un comptable.

Ces différences conceptuelles, conjuguées à certaines situations où le lieu de résidence est plus difficile à établir, peuvent donc avoir comme effet qu'une part de la population n'ait pas le même lieu de résidence au recensement et dans les données fiscales.

3. Données utilisées

Cette étude exploite les données d'un couplage d'enregistrements entre le recensement de 2011 et les données fiscales du fichier T1 sur la famille (T1FF) de 2010. Le T1FF est un fichier construit par Statistique Canada qui combine différentes sources fiscales, principalement les déclarations T1 des particuliers et les données de la Prestation fiscale canadienne pour enfants (PFCE), pour reproduire la population et les familles du pays. Cette banque de données a été retenue pour ce projet en raison de sa couverture très élevée de la population (95 %) et du fait qu'elle est souvent mobilisée par différents analystes et chercheurs, notamment pour le calcul des estimations officielles de la migration interne. Les données de l'Enquête nationale auprès des ménages (ENM) ont ensuite été ajoutées au couplage afin de profiter de la richesse des caractéristiques disponibles dans cette banque.

² La définition détaillée du concept du lieu habituel de résidence du recensement de 2011 est disponible ici : <http://www12.statcan.gc.ca/census-recensement/2011/ref/dict/pop126-fra.cfm> (page visitée le 5 novembre 2015).

³ Les déclarations fiscales incluent aussi la province ou le territoire de résidence au 31 décembre de l'année fiscale et au moment de remplir la déclaration.

Le couplage a été réalisé de façon déterministe au moyen de techniques éprouvées. Les personnes sont jumelées en cinq vagues successives à partir de leur nom, leur date de naissance, leur sexe et leurs informations familiales. En raison des objectifs de l'étude, l'information géographique n'a pas été utilisée. En effet, cette approche aurait biaisé la comparaison; les personnes couplées auraient été beaucoup plus susceptibles d'avoir le même lieu de résidence dans les deux sources.

Le fichier jumelé contient un peu plus de 18 millions de personnes, soit un taux de couplage de 57,0 % de la population du T1FF. De son côté, la partie du couplage réalisée avec l'ENM comprend environ 4 millions de personnes. Si le couplage comprend notamment un peu plus d'enfants et un peu moins de personnes âgées de 65 ans ou plus, ces écarts sont modérés de sorte que les données du jumelage demeurent fidèles aux données du T1FF et du recensement.

3.1 Définition du lieu de résidence pour cette étude

Dans le cadre de cette étude, le lieu de résidence est défini à partir du code postal. Cette décision repose à la fois sur le fait que les codes postaux représentent une unité géographique très fine qui se rapproche du lieu de résidence réel des individus, principalement dans les milieux urbains, et en raison des difficultés posées par l'harmonisation des adresses complètes des individus.

Les codes postaux du fichier couplé ont été nettoyés à partir d'une liste de codes postaux qui étaient valides en mai 2011. Selon ce critère, 95,9 % des individus appariés ont des codes postaux valides et ont donc pu être utilisés dans cette étude. Il y a correspondance du lieu de résidence si le code postal du T1FF est identique à celui du recensement.

3.2 Dates de référence des fichiers

Un élément important à retenir dans le cadre de la présente analyse est le fait que la date de référence du T1FF diffère de celle du recensement. Effectivement, si le recensement de 2011 a été administré le 10 mai 2011, les déclarations fiscales de l'année 2010 sont majoritairement remplies en mars ou en avril 2011. De plus, les renseignements géographiques envoyés à Statistique Canada par l'ARC aux fins de la construction du T1FF sont les informations les plus à jour disponibles à l'ARC au 31 décembre 2011. En conséquence, les données de l'année fiscale de 2010, recueillies au printemps de 2011, contiennent certaines données géographiques de la fin de l'année 2011 si le déclarant a changé de lieu de résidence après avoir rempli sa déclaration, a mis à jour ses renseignements et que l'ARC les a intégrés dans ses bases de données. Ce décalage peut donc influer sur la comparaison réalisée ici pour les individus qui ont déménagé.

4. Niveau de correspondance du code postal selon l'échelon géographique

Le tableau suivant rend compte du niveau de correspondance du lieu de résidence entre le recensement et le T1FF pour différents échelons géographiques.

Tableau 4.1-1
Taux de correspondance géographique (en %) selon différents échelons géographiques

| Échelons géographiques | Taux de correspondance |
|--------------------------------------|------------------------|
| Code postal | 92,9 |
| Région de tri d'acheminement | 95,5 |
| Subdivision de recensement | 93,9 |
| Division de recensement | 97,9 |
| Région métropolitaine de recensement | 98,7 |
| Province ou territoire | 99,7 |

Dans l'ensemble, 92,9 % des individus appariés ont le même code postal dans le T1FF et au recensement. Ce taux grimpe à 95,5 % en considérant les régions de tri d'acheminement (RTA)⁴ et à 93,9 % au niveau des subdivisions de recensement (SDR)⁵. Le niveau de correspondance un peu plus faible des SDR comparativement à celui des RTA peut s'expliquer par le fait que si les RTA constituent un échelon géographique très fin dans les milieux très urbains, elles forment des aires pouvant être beaucoup plus étendues que les SDR dans le reste du pays. Le niveau de correspondance continue d'augmenter pour les échelons géographiques plus agrégés alors qu'il s'approche de 98 % pour les divisions de recensement (DR)⁶ et dépasse 99 % à l'échelon provincial/territorial.

Ces résultats portent à croire que si le lieu de résidence inscrit dans les données fiscales peut différer de celui du recensement, cet écart n'est perceptible qu'à des échelons géographiques très fins. Cependant, les niveaux de correspondance relativement élevés observés ici ne doivent pas passer sous silence deux éléments. Premièrement, le fait que 92,9 % des personnes appariées ont le même code postal signifie que les codes postaux ne correspondent pas pour 7,1 % des personnes du couplage. Si les résultats du jumelage peuvent être inférés à l'ensemble de la population, près de 2,5 millions de personnes n'auraient pas indiqué le même code postal au recensement et dans les données fiscales. Il s'agirait d'un effectif équivalent à celui de la population de la région métropolitaine de recensement (RMR)⁷ de Vancouver, la troisième RMR la plus populeuse au pays.

Deuxièmement, il existe des variations régionales parfois importantes au chapitre du niveau de correspondance du code postal. La correspondance est notamment plus faible dans les territoires, où elle atteint un plancher de 85,1 % dans les Territoires du Nord-Ouest. Ce résultat signifie que moins de neuf habitants de ce territoire sur dix ont le même code postal dans le T1FF et au recensement. De plus, à des échelons locaux, le niveau de correspondance tend généralement à diminuer dans le cœur des zones urbaines et à être plus élevé dans les zones périphériques. La population qui réside au centre-ville de Montréal montre un niveau de correspondance des codes postaux inférieur à 85 %, soit le niveau le plus faible de toute la région de Montréal. Inversement, la population habitant plusieurs banlieues ainsi que certaines zones de l'ouest de l'île affiche régulièrement un niveau de correspondance supérieur à 95 %. Ces résultats émanent essentiellement de l'effet conjugué des différences régionales dans les caractéristiques de la population et des liens entre ces caractéristiques et les niveaux de correspondance des codes postaux.

5. Niveau de correspondance du code postal selon certaines caractéristiques

Le tableau suivant présente le taux de correspondance des codes postaux selon différentes caractéristiques⁸.

⁴ Une RTA correspond aux trois premiers caractères du code postal. En mai 2011, le Canada comptait 1 638 RTA.

⁵ Une SDR est un terme générique qui désigne les municipalités (telles que définies par les lois provinciales/territoriales) ou les territoires considérés comme étant des équivalents municipaux à des fins statistiques (p. ex., les réserves indiennes, les établissements indiens et les territoires non organisés).

⁶ Une DR est un groupe de municipalités voisines les unes des autres qui sont réunies pour des besoins de planification régionale et de gestion de services communs (comme les services de police et d'ambulance). Ces groupes sont créés selon les lois en vigueur dans certaines provinces du Canada.

⁷ Une RMR est un territoire formé d'une ou de plusieurs municipalités voisines les unes des autres qui sont situées autour d'un noyau. Une RMR doit avoir une population totale d'au moins 100 000 habitants et son noyau doit compter au moins 50 000 habitants.

⁸ Bien que les résultats présentés ici soient uniquement descriptifs, les liens entre ces caractéristiques et le niveau de correspondance des codes postaux demeurent statistiquement significatifs en tenant compte simultanément de l'effet de différents facteurs au moyen d'un modèle de régression logistique.

Tableau 5-1**Taux de correspondance des codes postaux (en %) selon différentes caractéristiques**

| Caractéristiques | Taux de correspondance |
|---|------------------------|
| Total | 92,9 |
| Groupe d'âge | |
| 0 à 9 ans | 91,5 |
| 10 à 19 ans | 93,6 |
| 20 à 24 ans | 91,0 |
| 25 à 29 ans | 89,2 |
| 30 à 39 ans | 92,0 |
| 40 à 49 ans | 94,7 |
| 50 à 59 ans | 95,4 |
| 60 à 69 ans | 94,9 |
| 70 à 79 ans | 93,1 |
| 80 ans et plus | 79,9 |
| A déménagé au cours de la dernière année * | |
| Non migrant | 95,4 |
| Migrant | 83,0 |
| Demeure dans un logement collectif | |
| Non | 93,4 |
| Oui | 31,4 |
| Ménage sur réserve * | |
| Non | 94,2 |
| Oui | 87,6 |
| Province/territoire de résidence et de travail † * | |
| Même province/territoire | 94,5 |
| Autre province/territoire | 89,4 |
| Présence du conjoint dans le ménage * | |
| Conjoint présent | 95,0 |
| Conjoint absent | 89,0 |
| Vague d'appariement | |
| 1) Appariement exact | 94,6 |
| 2) Contrainte moins sévère sur le nom | 93,3 |
| 3) Mix Match | 93,1 |
| 4) Au moins trois membres de la famille avec le même genre et date de naissance | 92,3 |
| 5) Personnes seules ajoutées | 84,8 |

* Caractéristiques tirées des individus appariés qui ont répondu à l'ENM. En raison des objectifs de l'étude et du taux de couplage, les résultats présentés ici qui sont tirés de cette enquête ne sont pas pondérés.

† Les personnes qui habitent dans la RMR d'Ottawa-Gatineau et dans les agglomérations de recensement (AR) de Campbellton, Hawkesbury et Lloydminster ont été exclues de cette analyse en raison du chevauchement de leur zone métropolitaine entre deux provinces.

Les données de ce tableau révèlent que plusieurs caractéristiques sont associées à des niveaux de correspondance plus bas des codes postaux du recensement et du T1FF. L'examen de la correspondance des codes postaux en fonction de l'âge révèle deux choses. Premièrement, le taux de correspondance fléchit modérément à partir du début de la vingtaine, chutant sous 90 %, pour ensuite amorcer une remontée vers l'âge de 30 ans. Ce résultat s'explique en partie par la propension plus élevée à déménager à ces âges. Effectivement, cette période du cycle de vie est généralement marquée par plusieurs déménagements motivés par le départ du foyer parental, la fréquentation d'établissements d'enseignement postsecondaire, l'insertion sur le marché du travail ou l'achat d'une première propriété de sorte que la propension à déménager atteint son apogée à ces âges (Dion et Coulombe, 2008). Or, le fait d'avoir déménagé au cours de l'année qui précède la tenue de l'ENM est également associé à un niveau de correspondance plus bas des codes postaux. Bien que les individus qui n'ont pas changé d'adresse montrent un niveau de correspondance de plus de 95 %, celle-ci chute sous 85 % pour ceux qui ont déménagé.

De façon corollaire, plusieurs caractéristiques associées à la mobilité, comme le statut d'immigrant ou le fait d'être locataire, sont aussi liées à un niveau de correspondance un peu plus bas. Si cette situation peut émaner du fait que les personnes qui déménagent soient plus susceptibles de garder des liens, du moins de manière temporaire, avec plus d'un lieu de résidence, il est manifeste qu'elle peut aussi provenir en large part du décalage dans les dates de référence des deux sources. Fait notable, même en tenant compte du fait d'avoir déménagé, les jeunes adultes affichent toujours un niveau de correspondance un peu plus bas, ce qui suggère que d'autres mécanismes que la mobilité pourraient expliquer cette situation.

Deuxièmement, la correspondance des codes postaux diminue énormément chez les personnes très âgées. En effet, elle diminue sous le seuil de 80 % à partir de 80 ans et atteint même près de 60 % chez les personnes âgées de 90 ans ou plus. Cette chute est attribuable à la forte propension de personnes âgées à vivre dans des logements collectifs. Or, le niveau de correspondance des codes postaux est spécialement faible pour les individus qui résident dans un logement

collectif (31,4 %). En se limitant aux personnes qui demeurent dans des logements privés, la correspondance des codes postaux des personnes âgées de 80 ans et plus se maintient d'ailleurs aux alentours de 90 %. Les personnes très âgées sont aussi plus susceptibles de faire remplir leur déclaration fiscale par un comptable ou par un membre de la famille. Ce facteur pourrait être associé à un niveau plus faible de correspondance des codes postaux du fait que le code postal inscrit sur la déclaration pourrait être celui de la personne qui a rempli la déclaration et non celui du déclarant.

Certaines caractéristiques associées à des situations où le lieu de résidence est plus difficile à bien établir sont également corrélées à des niveaux de correspondance un peu plus faibles. Ainsi, les individus qui ne travaillent pas dans leur province ou territoire de résidence et ceux qui ne demeurent pas avec leur conjoint montrent des niveaux de correspondance inférieurs à 90 %. En plus d'être plus mobiles que la moyenne nationale, ces individus sont aussi spécialement susceptibles de maintenir des liens avec plus d'un lieu de résidence de sorte qu'ils pourraient inscrire un code postal différent au recensement et dans les données fiscales. Les individus qui demeurent dans une réserve indienne montrent aussi un niveau de correspondance un peu plus faible (87,6 %). Les dynamiques économiques, sociales et politiques de ces zones sont parfois très différentes de celles du reste du pays.

Finalement, la correspondance des codes postaux suit aussi un gradient en fonction de la vague d'appariement. Les individus jumelés lors de la cinquième vague affichent un niveau de correspondance particulièrement bas (84,8 %). En raison de l'approche utilisée pour construire le couplage, les liens créés dans les dernières vagues sont considérés comme légèrement moins robustes de sorte que les chances d'avoir un faux appariement sont un peu plus grandes. Évidemment, dans cette situation, les codes postaux sont beaucoup moins susceptibles de correspondre. En outre, l'examen des personnes jumelées à cette vague révèle qu'elles sont plus mobiles et plus susceptibles de demeurer dans un logement collectif, deux caractéristiques aussi liées à des niveaux plus faibles de correspondance des codes postaux.

6. Conclusion

Le concept de lieu de résidence est très complexe. Les différences conceptuelles entre différentes banques de données sont susceptibles d'influer sur la cohérence des produits statistiques qui en découlent. Cette étude examinait la correspondance du lieu de résidence entre les données fiscales du T1FF et le recensement au moyen d'un couplage de données.

L'analyse a tout d'abord révélé que le niveau de correspondance du lieu de résidence est relativement élevé. En effet, 92,9 % des individus jumelés avaient le même code postal au recensement et dans le T1FF. Cette proportion grimpe jusqu'à plus de 99 % en considérant des échelons géographiques plus agrégés. Cet examen a également mis en exergue certaines variations, parfois importantes, dans le niveau de correspondance pour différentes franges de la population. Les jeunes adultes, les personnes très âgées, les personnes très mobiles et celles qui résident dans un logement collectif sont spécialement susceptibles d'afficher des taux de correspondance plus bas.

De manière plus générale, dans un contexte d'utilisation croissante de données administratives et de diversification des trajectoires sociales, les exercices de confrontation de données comme celui-ci gagneront vraisemblablement en importance non seulement pour mieux comprendre les fichiers disponibles mais aussi pour mieux apprécier les dynamiques démographiques qui caractérisent la population canadienne.

Bibliographie

Dion, P. et S. Coulombe (2008), « Portrait de la mobilité des Canadiens en 2006 : trajectoires et caractéristiques des migrants », *Rapport sur l'état de la population du Canada 2005 et 2006*, No 91-209-X au catalogue, p 83-134.

National Research Council (2006), *Once, Only Once, and in the Right Place: Residence Rules in the Decennial Census*. Washington, DC: The National Academies Press, 355 p.

Statistique Canada (2015), *Rapport technique du recensement : Couverture. Recensement de la population, 2011*, No 98-303-X au catalogue, 163 p.