
AU RECENSEMENT, 2,5 % DE DOUBLES COMPTES, D'APRÈS L'ÉCHANTILLON DÉMOGRAPHIQUE PERMANENT

Laurent TOULEMON(*), Sébastien DURIER(**), Benjamin MARTEAU(*)

(*) Institut national d'études démographiques (Ined)

(**) Insee, Direction des Statistiques Démographiques et Sociales

toulemon@ined.fr

sebastien.durier@insee.fr

benjamin.marteau@ined.fr

Mots-clés : enquêtes annuelles de recensement – précision – doubles comptes – Échantillon démographique permanent

Résumé

Depuis 2004, le recensement français a lieu par enquêtes annuelles. Cela a probablement conduit à une baisse des omissions, grâce à la professionnalisation des équipes en charge du recensement, devenu une opération quasi-continue, et à l'utilisation du répertoire informatisé des logements. Le nombre d'individus en double compte a, par contre, probablement augmenté. Sont en double compte potentiel les personnes qui vivent habituellement dans deux logements et sont susceptibles d'être recensées dans leurs deux logements. Si elles suivent les règles de collecte du recensement, elles ne sont recensées que dans un seul logement, et leur probabilité d'apparition dans une enquête annuelle est la même que celle des autres habitants ; si elles ne les suivent pas, elles donnent lieu à double compte : leur probabilité d'apparition dans les fichiers est double.

Les données du recensement rénové alimentent l'Échantillon démographique permanent (EDP) sans être apurées. Il en résulte que, depuis 2004, une personne recensée deux fois la même année, dans deux logements différents, renvoie deux « échos » et est présente deux fois dans l'EDP (bulletin double). En estimant la probabilité de bulletin double en cas de double compte, nous proposons une première estimation des doubles comptes au recensement, de 1,3 %. Cette estimation est fondée sur plus de 7 000 bulletins doubles et les variations aléatoires sont très faibles, l'intervalle de confiance étant de [1,24 ; 1,32] autour de l'estimation de 1,28 %. Mais deux facteurs limitent cette précision en pratique. Premièrement, certains « doubles » peuvent correspondre à deux personnes différentes. Une estimation pessimiste conduit à 15 % de « faux doubles », ramenant la proportion à 1,1 % ; inversement certains doubles ne sont peut-être pas appariés. L'autre facteur d'imprécision vient du fait que la méthode est basée sur l'observation des personnes en double compte dont les deux logements appartiennent au même groupe de rotation et sont donc recensés la même année. Mais rien ne garantit que la probabilité d'être en double-compte soit la même lorsque les deux logements d'un individu sont dans des groupes de rotation différents. Les personnes recensées une année donnée dans un de leurs logements, mais dont l'autre logement n'est pas recensé cette année-là, n'ont aucune raison de se savoir en double compte : pour elles le double compte est invisible.

À partir de 2011, une question sur la résidence un an auparavant a été introduite dans le recensement, en remplacement d'une question sur la résidence 5 ans auparavant. En se limitant aux personnes recensées dans un logement une année t et dans un autre logement l'année $t+1$, et déclarant en $t+1$ résider dans le logement depuis plus d'un an, on peut calculer une autre estimation des doubles comptes l'année t , de l'ordre de 2,6 % pour ceux dont les deux logements ne figurent pas

dans la même vague annuelle. Finalement, en supposant que la probabilité d'inclusion à tort d'un logement l'année t est de 2,6 % pour toutes les personnes dont un seul de leurs deux logements est sélectionné dans l'enquête annuelle de recensement une année donnée, nous estimons à 2,5 % la proportion des doubles comptes au recensement.

L'EDP permet également de décrire les situations familiales et professionnelles de ces habitants, dans leurs deux logements. Plus de la moitié sont enfants d'une famille et autant vivent hors famille dans un de leurs logements, tandis qu'un tiers sont adultes d'une famille. Un quart des personnes en double compte sont des enfants de parents séparés, qui partagent leur temps entre les deux logements parentaux. Les doubles comptes correspondent donc à des situations familiales spécifiques.

1. Introduction

Depuis 2004, le recensement est organisé de manière à ce que l'ensemble de la population résidant en France ne soit pas recensée au cours d'une même année, mais par des enquêtes annuelles de recensement (EAR). Cinq EAR successives sont ensuite fusionnées pour fournir les données du recensement de l'année médiane. Les taux de sondage annuel varient selon le type de commune dans lequel la personne est recensée : 20 % dans les communes de moins de 10 000 habitants, 8 % dans les communes de plus de 10000 habitants, 14 % en moyenne. En plus d'obtenir des informations sur les caractéristiques sociodémographiques des individus, le recensement a pour vocation de dénombrer les habitants. Des ajustements statistiques sont quelquefois nécessaires pour corriger les erreurs entraînant une surestimation ou une sous-estimation de la population. Les deux sources d'erreurs principales biaisant l'estimation de population sont les omissions d'individus, et les doubles comptes (individus comptés deux fois). Une enquête post-censitaire avait été réalisée pour le recensement de 1990 afin de déterminer la proportion d'omissions et de doubles comptes probables, mais aucune enquête comparable n'a eu lieu depuis, ni après le recensement de 1999 ni dans le cadre du recensement rénové. On ne dispose donc pas d'estimation de la qualité statistique du recensement et du comptage de la population qui en résulte.

Dans les grandes communes (de plus de 10 000 habitants), le recensement rénové se base sur le Répertoire des immeubles localisés (RIL), qui constitue une base de sondage des adresses, mise à jour régulièrement, dans laquelle les logements recensés chaque année sont tirés au sort. Dans les petites communes (de moins de 10000 habitants), l'ensemble de ses logements sont recensés l'année où la commune est échantillonnée. Le but de ce travail est de repérer les bulletins doubles (personnes effectivement recensées deux fois la même année) et d'en déduire la proportion de doubles comptes (personnes comptées deux fois au recensement, parce que leur probabilité d'être recensée est double) à partir des probabilités de sélection des logements. Cela est rendu possible grâce à l'Échantillon Démographique Permanent (EDP), une base de données juxtaposant de nombreuses sources statistiques pour un échantillon d'individus, notamment les bulletins individuels de recensement¹. Si deux bulletins d'une même EAR sont remplis pour un seul et même individu, les deux bulletins sont conservés dans la base étude de l'EDP, ce qui permet d'estimer le nombre de doubles comptes au recensement. Les deux bulletins des personnes recensées deux fois peuvent également être combinés pour fournir de l'information sur les deux logements où vivent ces personnes. La deuxième partie de ce travail s'attachera à déterminer les situations familiales et professionnelles des individus recensés deux fois.

2. Estimer les doubles comptes

2.1. Du recensement de 1990 au recensement rénové

À partir de l'enquête post-censitaire de 1990, la proportion de doubles comptes probables a pu être estimée à 0,7-0,9 %, celle des omissions à 2,7 %. Les omissions concernent principalement des logements occupés par une seule personne, qui sont moins facilement accessibles ou moins visibles. Les personnes rattachées à plusieurs logements ordinaires font partie des personnes à risque d'être soit omises soit en double compte. Les doubles comptes ou les omissions touchent majoritairement les jeunes adultes âgés de 20 à 24 ans (Coeffic, 1993).

Avec le recensement rénové, le premier recensement millésimé 2006 (regroupant les cinq vagues d'enquêtes annuelles de recensement 2004-2008) a donné lieu un ajustement des variations de population entre 1999 et 2006 de +661 000 individus (Pla, 2009). Le croisement entre données d'état civil et vagues d'enquête annuelle a donc conduit à revoir à la hausse l'estimation de la population courante (déduite de

¹ Les individus EDP sont sélectionnés via leur jour de naissance. Avant 2008, les individus nés les 4 premiers jours d'octobre sont automatiquement intégrés dans l'Échantillon démographique permanent dès qu'une information est disponible pour ces individus. Après 2008, les individus nés du 2 au 5 janvier, du 1^{er} au 4 avril, du 1^{er} au 4 juillet et du 1^{er} au 4 octobre sont également inclus dans l'EDP.

la population recensée en 1999 et des mouvements des années 1999 à 2005 – naissances, décès, entrées et sorties du territoire) pour l'ajuster à l'estimation issue du nouveau recensement. L'ajustement compense des changements dans les omissions les doubles comptes ; il a été réparti par sexe, âge et situation matrimoniale, mais ni selon d'autres variables (lieu de résidence, caractéristiques de la population) ni en variations des omissions et des doubles comptes.

Avec le recensement rénové, les omissions sont probablement moins nombreuses, le Répertoire des immeubles localisés permettant d'identifier l'ensemble des logements où le recensement doit avoir lieu chaque année dans les grandes communes. Les doubles comptes, quant à eux, sont probablement plus nombreux. L'exhaustivité du recensement ancien conduisait à identifier facilement les bulletins doubles. Si un bulletin est rempli dans chaque logement, les doubles comptes sont visibles pour les personnes et peuvent être évités par une règle simple consistant à remplir « un bulletin et un seul par personne ». Avec les enquêtes annuelles de recensement, le taux de sondage varie entre 8 % dans les grandes communes et 20 % dans les petites (Desplanques, 2008; Pan Ké Shon, 2007). Les doubles comptes ne conduisent alors au remplissage effectif de deux bulletins qu'environ une fois sur treize : il faut que les deux logements soient inclus dans la même vague de recensement. Les doubles comptes sont donc le plus souvent invisibles pour les personnes recensées : une personne susceptible de participer au recensement dans deux logements répondra au recensement si un seul de ses logements est sélectionné, sans se douter qu'elle est en double compte, son autre logement ne figurant le plus souvent pas dans la même vague annuelle de recensement.

2.2. Multirésidence et doubles comptes

Les situations de multirésidence ne devraient pas conduire à des doubles comptes au recensement : les personnes résidant habituellement dans deux logements doivent, si elles appliquent les règles d'inclusion et d'exclusion précisées dans la « feuille de logement » du recensement, remplir un bulletin dans un seul de leurs logements. L'énumération de la population est ainsi privilégiée par rapport à la description du contour des modes de vie de ces personnes multirésidentes (Coast et al., 2016).

Comme les habitants sont supposés être recensés dans un logement et un seul, les doubles comptes sont des erreurs de collecte rendus possibles par le fait que les personnes ont deux résidences « suffisamment habituelles » pour être considérées comme leur résidence principale. Une personne dans l'EDP fait l'objet d'un double compte si elle vit dans deux logements qui peuvent être inclus dans le recensement et si cette personne est prête à remplir un bulletin dans chacun de ses logements (Toulemon, 2018).

À différents âges de la vie, les situations familiales ou professionnelles peuvent aboutir au partage entre plusieurs domiciles, et donc à un risque d'être recensé à deux endroits différents une même année (Toulemon et Pennec, 2010). Le recensement prévoit des règles spécifiques pour ces situations, dans lesquelles on retrouve traditionnellement :

- Les enfants en garde alternée : ils doivent être recensés dans le logement du parent où ils passent le plus de temps. Si la garde est partagée de manière équitable, ils sont recensés dans l'endroit où ils ont passé la nuit de référence du recensement ;
- Les jeunes adultes qui vivent en semaine sur leur lieu d'études, dans des résidences universitaires, internats, chambre en ville, colocation... et qui rentrent le weekend chez leurs parents ou dans leur famille. La personne doit être recensée dans le domicile où elle passe la plupart de son temps, c'est-à-dire son logement sur son lieu d'études ;
- Les adultes ayant à la fois une résidence familiale et professionnelle doivent être recensés dans leur logement familial, même s'ils y passent moins de temps. La règle est donc opposée à celle pour les jeunes encore étudiants, rendant relativement complexe le remplissage des bulletins pour les habitants confrontés à ces situations ;
- Les adultes possédant une résidence secondaire et qui l'occupent à certains moments de l'année, effectuant ainsi des migrations saisonnières. La période de collecte du recensement (janvier-février) diminue néanmoins les risques pour cette population d'être comptée deux fois ;

- Les plus âgés, placés en institution, qui peuvent encore avoir un lien avec leur ancien logement ordinaire et y être recensés alors qu'ils passent désormais plus de temps en institution.

2.3. Comparaison avec d'autres pays

La plupart des pays utilisent le recensement pour dénombrer leur population. Aux États-Unis, des mesures ont été prises pour limiter les risques de double compte. Depuis 2010, le *Census Bureau* a ajouté une question de contrôle demandant pour chaque membre du ménage listé s'ils vivent habituellement ailleurs que dans le logement recensé. Cette question, plus directe à comprendre que les nombreuses règles indiquées dans les instructions de remplissage de la « feuille de logement » en France, permet d'identifier les logements où les individus n'auraient pas dû être recensés. Dans l'enquête américaine post-censitaire de 2010, l'estimation nette globale de la population n'est créditée que de 0,01 % de surestimation, mais la proportion de personnes comptée deux fois est estimée à 2,8 %. Elle est cependant moindre que celle du recensement de 2000 (Census Bureau, 2012).

En Angleterre, l'augmentation probable de la surestimation de la population lors du recensement de 2011 a été prévue à cause de l'augmentation de la proportion de personnes multirésidentes, et également à la suite d'un changement de méthodes de collecte dans le recensement (retours des informations par internet ou par la poste). Alors que la surestimation était estimée à 0,4 % en 2001, elle est estimée entre 0,5 et 1,0 % en 2011 (Abbott, 2009; ONS, 2012). En Suisse, dans le recensement de 2000, la surestimation est estimée à 0,35 % (Renaud, 2007).

Ces surestimations varient selon le groupe de population concerné. Les jeunes sont ainsi beaucoup plus souvent comptés deux fois que les autres groupes d'âge (Renaud, 2007), surtout lorsqu'ils sont étudiants (ONS, 2012). Les propriétaires de leur logement, plus âgés, font aussi partie de cette population à risque (Census Bureau, 2012).

3. Données et méthodes utilisées

3.1. L'Échantillon Démographique Permanent

L'3.1. L'Échantillon Démographique Permanent (EDP) est un échantillon panel d'individus, constitué par l'Insee depuis 1968. Pour ces individus sélectionnés à partir de leur date de naissance, plusieurs types de données sont collectés : état civil, recensement, données sociofiscales. Pour l'étude des doubles comptes, nous utilisons le recensement uniquement, dans sa version rénovée après 2004, qui n'a plus de caractère exhaustif, mais repose sur un taux de sondage défini en fonction du type de commune dans lequel il a lieu (Jugnot, 2014 ; Insee, 2018).

Bien que l'EDP se soit élargi dès 2008 à 16 jours de naissance, ce qui a conduit au quadruplement de l'échantillon, l'appariement automatique ainsi que le traitement manuel des litiges (plusieurs bulletins renvoyant à un individu EDP par exemple) ne se sont réalisés sur ces 16 jours qu'à partir de 2009, raison pour laquelle nous avons choisi de sélectionner les individus EDP nés au mois d'octobre en 2008 uniquement. Ce choix implique néanmoins un recalage sur les pondérations qui sont appliquées dans l'EAR, étant donné que ces poids sont recalés à partir des 16 jours EDP dès 2008.

L'appariement se fait sur la base du nom et du prénom de l'individu, réduisant ainsi les risques d'erreurs même si la couverture de la population EDP devant être recensée n'est pas complète. Les recensements à partir de 2004 ne sont plus nettoyés des bulletins doubles dans la base étude EDP, alors que pour les recensements précédents n'a été conservé qu'un seul bulletin individuel, même dans le cas où plusieurs entrées étaient appariées à un seul individu. Il est dès lors possible, au sein des recensements rénovés, de comparer les deux bulletins individuels de recensement appariés à un seul et même individu, d'analyser les caractéristiques de ces individus comptés deux fois, et d'en dégager une estimation de l'ampleur à l'échelle de l'ensemble des vagues de recensement.

Une note interne à l'Insee fait état d'un premier relevé des proportions d'individus avec bulletins multiples dans le recensement, par année (Mambetov, 2014). Ce comptage des bulletins doubles n'a pas donné lieu à une estimation de la proportion de doubles comptes, mais l'auteur y identifie des cas « douteux » où la

présence de deux bulletins ou plus pour le même individu EDP ne correspond pas certainement à un double compte. Nous considérons ici seulement les bulletins doubles qui correspondent aussi sûrement que possible à des doubles comptes au recensement. Pour cela, nous éliminons tous les cas douteux : les bulletins à « poids nul » (enfants mineurs recensés dans des ménages collectifs, enfants nés l'année du recensement), les individus présents plus de deux fois à une enquête de recensement, les individus présents deux fois dans le même logement. Nous avons finalement conservé les bulletins des jumeaux EDP, après avoir vérifié que la gémellité n'entraîne pas de « faux doubles » quand les deux jumeaux sont recensés la même année. Par ailleurs, nous sommes revenus au fichier de saisie pour récupérer une information suffisamment précise (et non imputée) sur les adresses pour distinguer, dans les grandes communes, les logements différents à la même adresse ou à des adresses différentes, et sur l'ancienneté de résidence dans les logements, utilisée dans l'analyse de deux EAR successives.

Les résultats proviennent de la Base étude 2015, reprenant les vagues de recensements de 2004 à 2015.

3.2. Probabilités d'être compté deux fois dans le recensement rénové

Appelons « bulletin double » la présence effective de deux bulletins de recensement pour le même individu², et « double compte » le fait de résider dans deux logements et d'être susceptible de remplir (ou de faire remplir par une autre personne) un bulletin individuel de recensement dans chacun de ces logements. On peut calculer la probabilité de double compte d'un individu possédant deux bulletins de recensement, si l'on connaît les probabilités d'inclusion des deux logements et leur corrélation. On en déduit alors la proportion de doubles comptes dans une vague de recensement dans un premier temps.

Probabilités théoriques

Une personne a deux logements A et B dans lesquels elle peut être recensée³. Supposons dans un premier temps que les deux logements sont indépendants avec chaque année la probabilité de tirage de p (sans remise entre t et $t+1$).

En appelant A, B, AB et O respectivement les événements correspondant aux logements « A est recensé mais pas B », « B est recensé mais pas A », « A et B sont recensés » et « Ni A ni B ne sont recensés », et R, R1, R2 les événements « l'individu est présent au recensement », « l'individu est présent une fois au recensement » et « l'individu est présent deux fois au recensement », on a :

$$P(A) = p(1-p) ; P(B) = p(1-p) ; P(AB) = p^2 ; P(O) = (1-p)^2$$

L'individu peut être recensé dans A ou dans B :

$$P(R) = P(A \text{ ou } B) = P(A)+P(B)-P(AB) = 2p - p^2 = 2p(1-p) + p^2$$

$$P(R1) = 2 p(1-p)$$

$$P(R2) = p^2$$

D'où la probabilité de bulletin double en cas de double compte : $p(R2|R) = p / (2-p)$.

En appelant N la population totale, S les personnes ayant un seul logement, DC le nombre de personnes donnant lieu à un double compte dans la population, n le nombre de bulletins au recensement et bd le nombre de paires de bulletins doubles au recensement, on a :

$$N = S + DC$$

² Pour rentrer dans notre champ, les deux bulletins doivent être de poids non nuls et apparaître dans deux logements différents.

³ On peut fixer en toute généralité que la personne serait recensée « à raison » en A et « à tort » dans B d'après les règles du recensement, mais sans supposer qu'elle le sache. Cela permet de compter les personnes recensées à tort. Nous n'avons pas cherché à faire la distinction ici : nous supposons que toutes les personnes en double compte ont deux logements, l'un où elles peuvent être recensées à raison et l'autre où elles le seraient à tort. Il peut y avoir des personnes comptées à tort sans être en double compte (par exemple si elles ont leur logement principal à l'étranger).

$$DC = bd / p^2$$

$$N = (n/p) - DC$$

$$DC / N = (1 / p) (bd / (n-bd/p)) \approx (1 / p) (bd / n) \text{ en négligeant } bd/p \text{ devant } n.$$

Avec p de l'ordre de 14 %, on obtient $p(R2|R) = 7,5 \% = 1/13$. Cela signifie que, pour une personne ayant deux logements et présente au recensement dans un de ses logements, la probabilité que les deux logements participent au recensement la même année est très faible. En supposant que les personnes recensées la même année dans deux logements pourraient se rendre compte du double compte (et donc finalement ne remplir qu'un bulletin), cela ne concerne qu'une toute petite part des doubles comptes. Sous l'hypothèse que le double compte est invisible pour les personnes recensées deux années différentes, l'estimation des doubles comptes à partir des personnes recensées dans un logement une année donnée et dans un autre logement une autre année permettra de mieux mesurer les doubles comptes. Autre conséquence, nous allons estimer les doubles comptes en utilisant un facteur multiplicatif inverse du taux de sondage du recensement. Ce facteur multiplicatif, similaire à celui utilisé dans les méthodes de capture – recapture, limite la précision de nos estimations.

Estimation des probabilités pratiques

Le taux de sondage des enquêtes de recensement est plus complexe. Comme dit précédemment, le taux de sondage est différent dans les « grandes communes » (plus de 10 000 habitants) et dans les « petites communes » (moins de 10 000 habitants). La moitié des habitants vivent dans une petite commune, l'autre moitié dans une grande commune. Chaque EAR contient un cinquième (20 %) des petites communes, qui sont recensées exhaustivement. Les grandes communes sont incluses dans chaque EAR, mais ne sont pas recensées de manière exhaustive. Chaque année, 8 % des individus de ces grandes communes sont recensés. Une base de sondage des adresses permet de déterminer trois strates de tirage : 1) les grandes adresses : un seuil de logements minimum (qui change chaque année) définit les grandes adresses ; 2) les adresses nouvelles : nouvelles constructions ou rénovations conduisant à une modification du nombre de logements à l'adresse ; 3) les petites adresses connues : toutes les autres adresses.

Les grandes adresses et les adresses nouvelles sont tirées exhaustivement, les petites adresses sont échantillonnées pour obtenir 40 % des logements du groupe d'adresses (Insee, 2012). Chaque année un cinquième de ces adresses sont retenues, les cinquièmes étant aussi semblables et dispersés que possible. Dans les adresses retenues tous les logements sont recensés. Les informations contenues dans la base EDP permettent de distinguer les logements situés dans une petite commune et dans une grande commune, ainsi que, dans les grandes communes, les logements situés à la même adresse⁴.

Les probabilités de sélection simultanée de deux logements sont donc en théorie les plus élevées pour un individu ayant ses deux logements dans la même petite commune (puisque dans la petite commune recensée en t tous les logements sont recensés) (Tableau 1). Viennent ensuite celles de logements situés dans la même grande commune (et plus précisément à la même adresse), de logements dans deux petites communes différentes, de logements localisés dans une petite et une grande commune et enfin dans deux grandes communes (adresses) différentes.

⁴Nous sommes reconnaissants à Pascal Ardilly de nous avoir fait parvenir deux notes sur l'EDP, l'une sur les pondérations de l'EDP à des dates t et $t+5$, et l'autre sur les doubles comptes. Ses résultats sont conformes aux nôtres, la première pouvant être appliquée aux personnes recensées en t et $t+5$, mais également aux personnes recensées dans leurs deux logements en t (Favre 2018). Ces notes datent de 2015, mais nous n'en avons eu connaissance qu'après avoir réalisé les estimations présentées ici.

Tableau 1 : Probabilité de sélection de deux logements pour une EAR

Premier logement	Deuxième logement				
	Petite commune		Grande commune		
	La même	Une autre	Même adresse	Autre adresse	Autre commune
Petite commune	1/5 * 1	1/5 * 1/5			1/5 * 1/12,5
Grande commune		1/12,5 * 1/5	1/12,5 * 1	1/12,5 * 1/12,5	1/12,5 * 1/12,5

Source : Toulemon, 2018.

Pour des personnes en double compte, la localisation du deuxième logement est nécessairement indépendante de la construction des sous-échantillons, pour chacune des catégories présentées au tableau 1, étant donné que les personnes ne savent pas comment sont construits les groupes de rotation. On suppose donc que les logements sont indépendants et que la probabilité d'être recensé deux fois est le produit des probabilités d'inclusion dans chaque logement ; on calcule le poids des personnes recensées deux fois la même année comme l'inverse de leur probabilité d'être recensées deux fois (Tableau 2).

Tableau 2 : Poids des individus en double compte selon la probabilité de sélection des deux logements

Premier logement	Deuxième logement				
	Petite commune		Grande commune		
	La même	Une autre	Même adresse	Autre adresse	Autre commune
Petite commune A	5	25			62,5
Grande commune C		62,5	12,5	156,25	156,25

Source : Toulemon, 2018.

Une fois les individus avec bulletins doubles pondérés selon les règles ci-dessus, ceux-ci sont rapportés à l'ensemble des individus pondérés. Pour éviter les « faux bulletins doubles », nous avons éliminé les personnes apparaissant plus de deux fois la même année, ou étant recensés deux fois dans le même logement. Une vérification des modalités pratiques des procédures d'attribution des documents aux individus EDP indique que ce risque est extrêmement faible, les cas ambigus ou problématiques lors de l'appariement étant renvoyés à une validation humaine. Pour le confirmer nous avons calculé un score de proximité entre les deux bulletins associés à la même personne.

Bulletins doubles en t et t+1

Une personne en double compte en t peut également donner lieu à un bulletin double en t et une autre année, mais ces bulletins doubles sont plus délicats à repérer. Depuis 2011, le bulletin individuel recensement inclut une question sur la résidence un an auparavant : « Où habitiez-vous le 1^{er} janvier [de l'année dernière] ? », tandis que pour les vagues de recensement avant 2011 la question portait sur la résidence cinq ans auparavant. Sous certaines réserves que nous préciserons ci-dessous, on peut donc également estimer les doubles comptes en t à partir des bulletins doubles en t et t+1. D'un point de vue théorique, on peut calculer la probabilité qu'une personne ayant deux logements A et B en t (en l'absence de changement de résidences, la personne réside dans les mêmes logements en t+1) soit recensée en t et le soit également l'année suivante, en t+1 :

L'individu peut être recensé dans A ou dans B en t. S'il est recensé une seule fois, on ne l'observe qu'en t :

$$P(R) = P(A \text{ ou } B) = P(A) + P(B) - P(AB) = 2p - p^2 = 2p(1-p) + p^2$$

$$P(R1) = 2p(1-p)$$

S'il est recensé en t et en t+1, ce peut être soit en A puis B, soit en B puis A :

$$P(R2) = 2p^2$$

D'où la probabilité de bulletin double (un en t et un en t+1) en cas de double compte :

$$p(R2|R) = 2p / (2-p)$$

Si la personne a quitté un de ses logements entre t et $t+1$, on ne peut la retrouver que dans l'autre :

$$P(R2) = p^2 \text{ et } p(R2|R) = p / (2-p)$$

Si la personne a quitté ses deux logements entre t et $t+1$, on ne la retrouve pas :

$$P(R2) = 0 \text{ et } p(R2|R) = 0$$

On notera qu'une personne en double compte en t a deux fois plus de chances d'être recensée deux années de suite que d'être recensée deux fois la même année, si elle n'a pas déménagé. Avec p de l'ordre de 14 %, on obtient $p(R2|R) = 15 \% = 2/13$. La probabilité d'un bulletin double en t et $t+1$ permet de repérer les doubles comptes en t à partir de la même logique qu'à partir des bulletins doubles en t . Cinq différences importantes méritent cependant d'être notées.

Premièrement, le champ n'est pas tout à fait complet : dans les situations où les deux logements font partie de la même petite commune ou figurent à la même adresse d'une grande commune, un bulletin double est impossible. Deuxièmement, nous avons également vérifié que les changements des définitions d'adresse (petite ou grande adresse, communauté ou ménage ordinaire) ainsi que les changements de taille de commune (passant la barre des 10 000 habitants) ou les fusions de communes n'entraînaient pas des « faux bulletins doubles »⁵. Troisièmement, le double compte en t n'est observable en t et $t+1$ que si la personne, recensée dans un logement A en t , n'a pas déménagé de son logement B entre t et $t+1$, et a bien indiqué en $t+1$ qu'elle résidait déjà en B un an auparavant (c'est-à-dire en t). Nous avons donc restreint le champ des bulletins doubles en t à un champ comparable, en termes d'ancienneté dans les deux logements. Nous avons inclus les personnes résidant dans leurs deux logements depuis plus d'un an, divisé par deux la pondération de celles qui étaient dans un de leurs logements depuis moins d'un an et dans l'autre depuis plus d'un an, et exclu celles qui étaient dans leurs deux logements depuis moins d'un an⁶.

Les deux dernières différences sont plus critiques. Quatrième différence, le risque de « faux bulletins doubles » est beaucoup plus important, parce que toutes les personnes qui ont déménagé peuvent être incluses deux années de suite dans le recensement : à condition de changer de petite commune ou de changer d'adresse dans une grande commune, une personne recensée en t dans son ancien logement a 14 % de chances d'être à nouveau interrogée dans son nouveau logement (et plus encore si ce nouveau logement est un logement neuf ou une grande adresse). Cela ne donne pas lieu à double compte : d'une année à l'autre, on peut considérer le tirage des EAR comme un tirage « avec remise » pour les personnes qui ont déménagé. Parmi les personnes recensées lors de deux vagues successives et $t+1$, il faut donc distinguer très précisément celles qui ont déménagé entre t et $t+1$ et celles qui résidaient déjà en t dans le logement où elles sont recensées en $t+1$. Comme les déménagements sont nombreux dans la population, il faut bien veiller à ne pas considérer des personnes résidant dans leur logement en $t+1$ depuis moins d'un an comme en double compte en t . Nous avons donc utilisé la variable de saisie sur la résidence un an auparavant, non imputée, et exclu toutes les situations où la double résidence est douteuse.

Enfin, cinquième différence, nous supposons que les doubles comptes en t sont mieux mesurés par les bulletins doubles en t et $t+1$. On peut supposer qu'une personne en double compte, si elle est recensée la même année dans ses deux logements, risque de s'en apercevoir et de ne pas remplir deux bulletins. À l'inverse, si un seul logement figure dans l'échantillon d'une enquête annuelle, la personne n'a pas de raison de se soucier du double compte, et risque même de ne pas s'en soucier. On suppose donc que les

⁵Nous remercions Valérie Roux et Gwennaël Solard, du département de la Démographie de l'Insee, de nous avoir alertés sur ces risques. Ces cas se montent au total à moins de 60 individus EDP, sur un total de 7 969 individus en bulletin double sur deux années sur la période 2011-2015.

⁶ Les champs sont comparables sous l'hypothèse que la proportion de ceux qui seront encore dans leur logement 1 an après le recensement est la même que ceux qui sont dans leur logement depuis plus d'un an. En régime stationnaire, si la proportion de personnes en doubles comptes qui abandonnent un de leurs logements est constante, cette hypothèse est vérifiée.

doubles comptes seront plus nombreux à partir d'une estimation sur deux années différentes que sur une seule année. Nous n'avons pas regardé les bulletins doubles après plus d'un an, parce que la question sur la résidence porte sur « un an avant » (la variable d'ancienneté dans le logement n'est pas individuelle) ; d'autre part les déménagements deviennent trop nombreux et risquent de perturber la mesure. Sous l'hypothèse que quelques doubles comptes en t et t+1 seraient visibles par les personnes et donc évités, l'estimation des doubles comptes en serait sous-estimée. Rappelons qu'une personne en double compte participant au recensement a théoriquement des probabilités 1/13 et 2/13 que son autre logement soit inclus respectivement la même année ou l'année suivante, soit au total moins d'une chance sur quatre.

4. Résultats

4.1. Proportion de bulletins doubles et estimation d'individus en doubles comptes

Chaque année entre 2004 et 2008, environ 460 bulletins sont des « doubles » et 230 individus EDP apparaissent 2 fois dans la même vague d'EAR. En 2009, avec le passage de 4 à 16 jours, le nombre de bulletins doubles est lui aussi multiplié par 4, et ce sont près de 2 000 personnes pour lesquelles chaque année le recensement « renvoie deux échos ». Une fois éliminés les cas douteux, ce ne sont plus que 208 puis 940 individus EDP qui apparaissent deux fois exactement dans deux logements différents chaque année, 7 597 pour l'ensemble des années 2004-2015 (tableau 3). Ramené à l'effectif total de l'EDP, la proportion de bulletins doubles est stable : 0,5 % de bulletins doubles, et 0,24 % des individus EDP qui apparaissent deux fois. Cela devrait donc conduire à une estimation des doubles comptes de l'ordre de $0,24\% / 14\% = 1,7\%$ de personnes en double compte au recensement, qu'elles soient recensées effectivement dans l'un ou l'autre de leurs logements, ou dans les deux (tableau 3).

Tableau 3 : Nombre d'individus EDP concernés par des bulletins doubles, par année de recensement

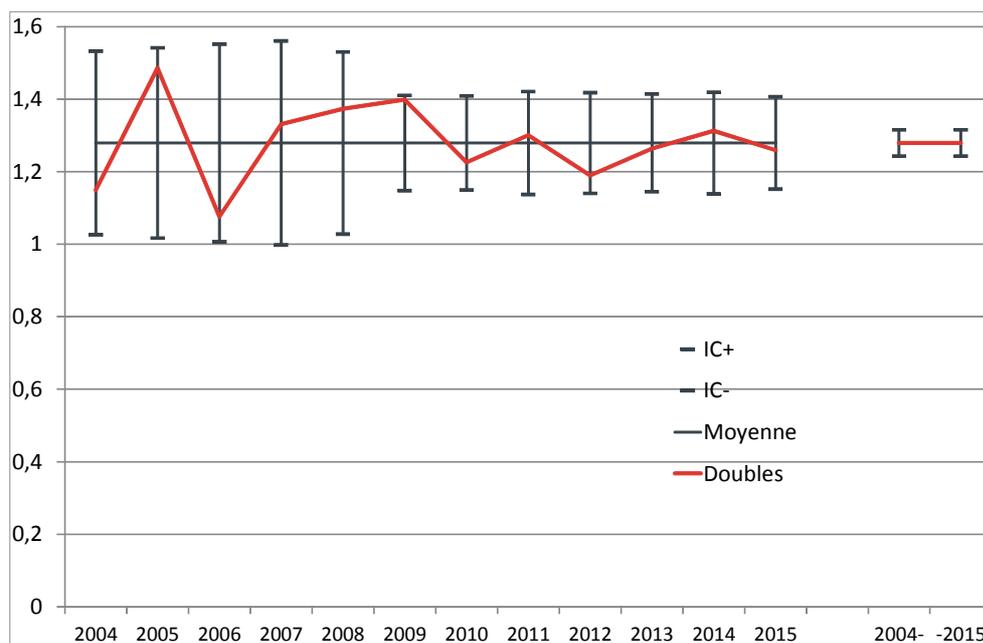
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	Total
Nombre d'individus avec BI doubles (non pondéré)	195	215	202	209	222	1025	886	882	867	873	911	1110	7597
Nombre d'individus recensés (non pondéré)	91625	93704	94666	94803	95703	387763	389549	389439	387978	392034	391805	393004	3202073
% d'ind. avec BI doubles par rapport au nombre de BI	0,213	0,229	0,213	0,220	0,232	0,264	0,227	0,226	0,223	0,223	0,232	0,282	0,237
Effectif pondéré d'individus en double compte	7437	9951	7201	9009	9374	38261	33606	35877	32738	34994	36592	34696	289735
% d'individus en double compte (pondéré)	1,149	1,486	1,076	1,331	1,374	1,399	1,226	1,300	1,190	1,263	1,313	1,260	1,280

Champ : BI des individus nés un jour EDP, à l'exclusion des BI avec un poids nul. EAR 2004-2015.

Source : Base étude 2015 de l'EDP. Les quelques individus présents plus de deux fois sont exclus de ces totaux.

Cependant, une fois les individus pondérés avec leur poids au recensement, la proportion de doubles comptes (rapportée à la population recensée) s'élève à 1,28 %. Nous verrons que cela s'explique par le fait que le poids moyen des personnes recensées deux fois est plus faible, car ils vivent souvent dans des logements neufs ou des grandes adresses. Nous avons calculé des intervalles de confiance des estimations annuelles (tenant compte de la variance des poids), pour valider la précision de notre estimation et pour tester la présence d'une tendance temporelle. Les variations aléatoires sont faibles : l'intervalle de confiance est de +/- 0,047 point, ce qui en erreur relative se chiffre à +/- 4 % (l'erreur relative est de l'ordre de 20 % pour une seule vague d'enquête avant 2008) (Figure 1). L'année 2009 apparaît comme particulière puisque la proportion d'individus comptés deux fois approche la limite supérieure de l'intervalle de confiance. Première année où l'appariement automatique pour les 16 jours EDP est réalisé, les erreurs d'appariement pourraient expliquer cette proportion plus élevée, qui descend de manière stable assez vite en dessous de la moyenne pour les 5 vagues de recensement à partir de 2010.

Figure 1 : Proportion d'individus en doubles comptes par année de recensement



Champ : BI des individus nés un jour EDP, à l'exclusion des BI avec un poids nul. EAR 2004-2015.

Source : Base étude 2015 de l'EDP.

Lecture : En 2004, 1,15 % des individus recensés sont en double compte. L'intervalle de confiance est de [1,01 %-1,53 %] autour de la valeur moyenne de 1,28 %, au seuil de 5% bilatéral.

4.2. Pondération des bulletins doubles selon le type de commune

En pratique, les pondérations observées sont différentes et inférieures aux pondérations théoriques (Tableau 4). Pour les bulletins simples, le poids moyen est de 7 (12,4 dans les grandes communes, 5 dans les petites communes), conformément au plan de sondage. Les petites communes sont surreprésentées dans le recensement (72 % non pondérés, ramenés à 52 % une fois pondéré).

Les doubles comptes sont plus fréquents dans les grandes communes et, dans les grandes communes, dans les grandes adresses et les logements neufs, ce qui aboutit à un poids moyen observé pour les personnes en bulletin double plus faible, de l'ordre de 36 (soit un taux de sondage de $1/6^2$) au lieu de 50 (pour un taux de sondage de 14 %, environ $1/7$), et donc une estimation plus basse, qu'anticipé à partir des calculs théoriques et de la proportion de bulletins doubles.

Tableau 4 : Répartition des poids par type de commune et de logement

	Effectifs non pondérés	Répartition en %	Poids théorique	Poids observé	Poids obs. / poids th.	Répartition pondérée (%)
Ensemble	3 202 073	100,00	7,09	7,07	1,00	100,00
Recensés une fois	3 194 476	99,76	7,00	7,00	1,00	98,81
Petite commune	2 332 109	72,82	5,00	5,00	1,00	51,51
Grande commune	862 367	26,94	12,50	12,42	0,99	47,30
Recensés deux fois	7 597	0,24	45,52	35,56	0,81	1,28
Petite c., grande c.	2 154	0,07	62,50	51,23	0,82	0,49
Deux grandes c.	649	0,02	156,25	113,77	0,75	0,33
Deux petites c.	2 450	0,08	25,00	25,11	1,01	0,27
Grande, ad. diff.	282	0,01	156,25	113,75	0,73	0,14
Même petite c.	1 804	0,06	5,00	5,00	1,00	0,04
Grande, même ad.	258	0,01	12,50	11,44	0,91	0,01

Champ : BI des individus nés un jour EDP, à l'exclusion des BI avec un poids nul. EAR 2004-2015. c. : commune ; ad. : adresse ; diff. : différentes.

Source : Base étude 2015 de l'EDP.

Note : Les individus recensés une fois sont pondérés à partir de la variable poids_ea. Les individus recensés deux fois sont pondérés par l'inverse de la probabilité de sélection simultanée des deux logements.

D'autres sources d'incertitudes pouvant faire varier notre estimation sont possibles :

- Des appariements erronés, qui conduisent à une surestimation des doubles comptes. Une vérification des modalités pratiques des procédures d'attribution des documents aux individus EDP semble indiquer que ce risque est extrêmement faible. Pour le valider nous calculons ci-dessous un score de proximité entre les deux bulletins associés à la même personne.
- Le fait de récupérer deux bulletins de recensement pour une personne EDP, mais dont un seul est apparié à la bonne personne EDP (des faux négatifs, également possibles mais impossibles à identifier).
- Surtout, les bulletins doubles sont peut-être rares car quand les deux logements d'une personne en double compte sont recensés la même année *t* le double compte est visible, et les règles cherchant à attribuer chaque habitant à un logement et un seul sont plus probablement respectées. À l'inverse, une personne habitant dans deux logements dont un seul est inclus dans la vague de l'année *t* risque de ne pas se poser la question de son inclusion éventuelle également dans son autre logement, et de participer au recensement sans percevoir qu'il est en double compte (et donc peut-être recensé à tort). Nous estimons plus loin cette probabilité à partir des personnes recensées deux années consécutives.

4.3. Création d'un score d'appariement pour discerner la probabilité de vrais doubles comptes

Les erreurs d'appariement peuvent survenir et conduire à attribuer deux bulletins de personnes différentes à un seul individu EDP. La proportion d'erreurs d'appariement n'est pas connue. On peut imaginer, comme cela a déjà été documenté, que les individus EDP nés à l'étranger sont plus soumis à ces erreurs d'appariement. En effet, la gestion automatique qui apparie les bulletins de recensement a un pourcentage de réussite plus faible pour ces individus (Jugnot, 2014). Nous avons vérifié que ces difficultés d'appariement ne sont pas beaucoup plus fréquentes pour les individus en bulletin double. En moyenne, 17 % des identifications passent par une opération manuelle après échec de l'appariement automatique ; ce taux atteint 20,5 % pour les identifications aboutissant à un double compte simultané et 19 % pour les doubles comptes repérés à partir de deux bulletins remplis deux années consécutives.

Afin de déterminer à quel point on peut être sûr que deux bulletins remplis peuvent concerner une seule et même personne, plusieurs variables ont été utilisées pour créer un « score d'appariement » entre deux bulletins. Les variables choisies, non redressées dans le recensement, sont (autant que possible) indépendantes des lieux de résidence des individus au moment du recensement. Ainsi, le score ne dépend pas des réponses qui pourraient varier selon l'endroit dans lequel l'individu est recensé (si c'est effectivement le même individu).

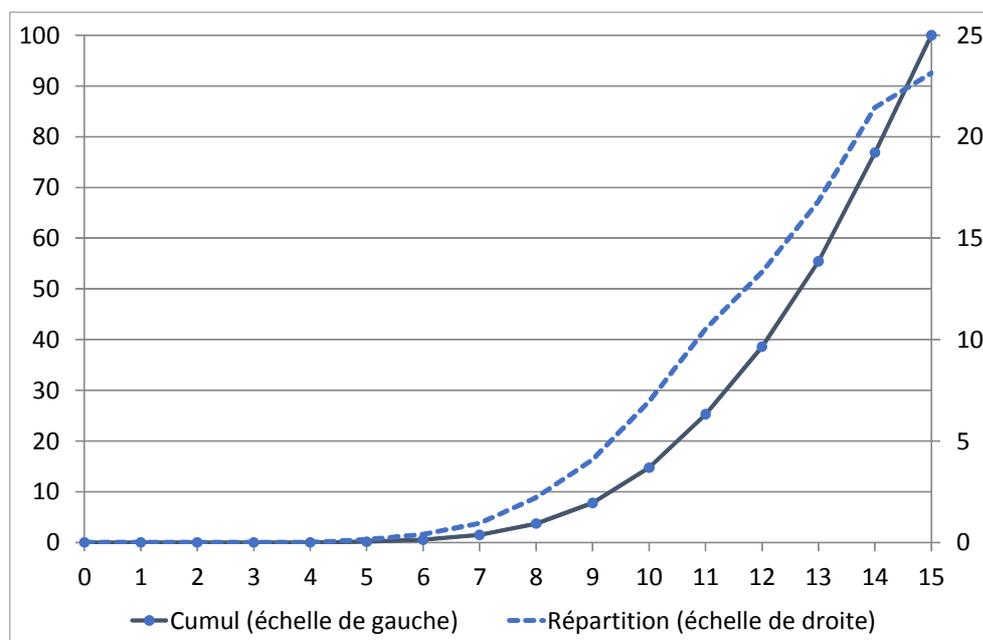
Si deux bulletins individuels ne contiennent pas les mêmes informations sociodémographiques de base, tel que l'année de naissance, le lieu de naissance ou encore le sexe, on peut s'attendre à des erreurs d'identification et des « faux » doubles comptes.

15 variables ont été sélectionnées pour la création du score :

- Commune de naissance
- Date de naissance (trois variables : année, mois, jour de naissance)
- Sexe
- Nationalité
- Diplôme
- État matrimonial
- Année d'arrivée en France
- Type d'activité
- Inscription dans un établissement d'enseignement
- Commune du lieu de travail
- Commune du lieu d'études
- Catégorie socioprofessionnelle
- Vie en couple (cette variable peut cependant varier si l'individu vit en couple dans un seul de ses deux logements).

Les graphiques ci-dessous reprennent la répartition des individus selon le score obtenu. La moyenne pondérée est de 12,75 modalités identiques pour l'ensemble des EAR.

Figure 2 : Répartition du nombre de variables identiques entre les deux bulletins de recensement (score d'appariement pour 15 variables mesurées) et fréquence cumulée de réponses identiques (en %)



Champ : individus nés un jour EDP, recensés deux fois la même année (bulletins doubles). N=7 597.

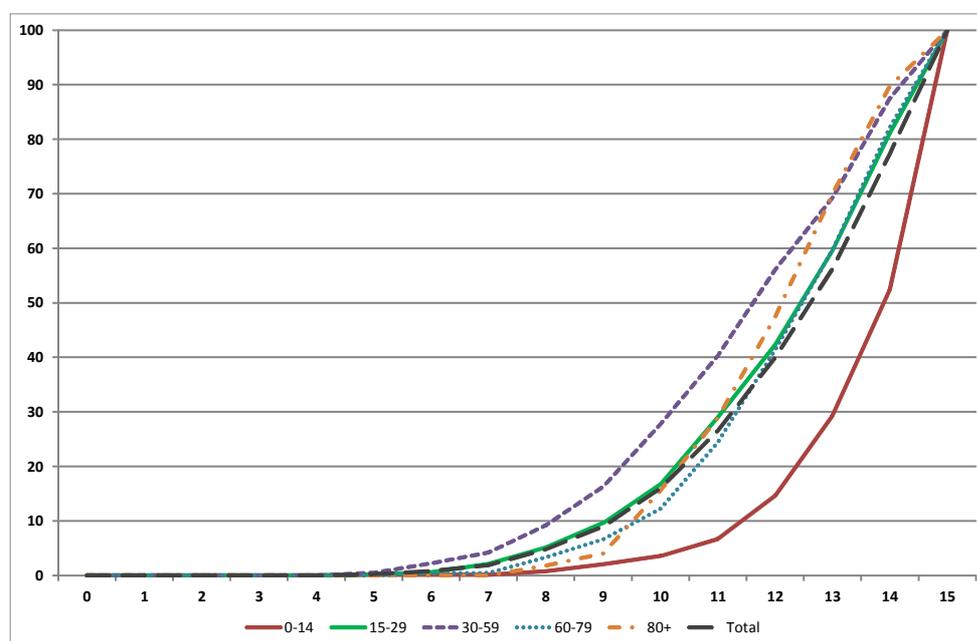
Source : Base étude 2015 de l'EDP.

Lecture : Parmi l'ensemble des individus recensés deux fois lors d'une année de recensement, 16 % ont 10 informations identiques ou moins entre leurs deux bulletins de recensement remplis, et 7 % ont exactement 10 réponses identiques.

Certaines variables sélectionnées n'ont pas beaucoup d'impact sur le nombre total d'informations concordantes : c'est le cas du sexe et de la date de naissance. D'autres informations par contre, comme la catégorie sociale ou le niveau de diplôme, sont beaucoup plus corrélées au score final obtenu pour chaque individu. Ces variables sont plus propices à identifier les « faux » doubles comptes présents dans notre base de données.

Au total, sur les 15 variables choisies, moins de 1 % des observations ont 6 informations ou moins en commun, indiquant que très peu d'individus possèdent des caractéristiques éloignées selon les deux bulletins de recensement (Figure 2). 15,5 % des individus ont 11 modalités ou moins identiques. 43,3 % des individus ont 14 ou 15 informations identiques (soit le maximum), avec une nette surreprésentation des enfants dans ces catégories, puisque ceux-ci ont de nombreuses informations manquantes dans leur bulletin qui, mécaniquement, sont similaires d'un bulletin à l'autre (Figure 3).

Figure 3 : Score d'appariement par groupe d'âge (2004-2015)



Champ et source : voir figure 2.

Les informations présentes dans les deux bulletins peuvent différer même si les deux bulletins concernent le même individu. Les lieux de travail, catégorie socioprofessionnelle, vie en couple, peuvent différer d'un logement à l'autre, si les bulletins sont remplis l'un par la personne elle-même, l'autre par un proche (enfant, parent, conjoint...). Cependant la probabilité d'erreur d'appariement (et donc de faux double compte) augmente au fur et à mesure que le score diminue.

En utilisant notre score (divisé par 15) comme une probabilité de « vrai bulletin double », on obtient une probabilité moyenne de 85 % et une estimation basse de la proportion de double compte de 1,09 %.

4.4. Comparaison avec deux vagues successives de recensement t ; $t+1$

On considère les personnes recensées en t dans un logement et recensées à nouveau dans un autre logement en $t+1$, et qui déclarent ne pas avoir changé de logement la deuxième année d'observation : elles disent en $t+1$ vivre dans le même logement depuis un an ou davantage, or leur lieu de résidence comme on le retrouve dans le bulletin en t ne correspond pas à leur lieu de résidence en $t+1$. On en déduit alors qu'elles résidaient dans deux logements en t et étaient alors susceptibles d'être comptés deux fois. Ces personnes sont comparées aux individus comptés en double une année t , déclarant ne pas avoir

déménagé dans aucun de leurs deux logements ou déménagé dans un seul de leurs deux logements (dont on ne compte que la moitié d'entre eux pour la comparaison). L'univers comparé est donc le même.

Pour les années de recensement qui s'étendent de 2011 à 2015, la proportion de doubles comptes atteint 1,27 % (Figure 1). En enlevant les personnes dont les deux logements sont dans la même petite commune ou à la même adresse dans une grande commune, la proportion atteint 1,21 %. En supprimant les personnes résidant dans leurs logements depuis moins d'un an et la moitié de celles résidant dans un de leurs logements depuis moins d'un an (et depuis plus d'un an dans l'autre), la proportion descend à 0,76 %, ce qui est conforme à ce que l'on sait des épisodes de multi-résidences dans les enquêtes longitudinales : d'après l'enquête SRCV, la part des multi-résidences durant plus d'une année étant de l'ordre de 50 % à 65 % (Toulemon et Pennec, 2009). Pour le même champ, l'estimation directe est de 1,60 % pour deux vagues successives, ce qui conduit à une estimation de 2,54 % pour les cas où de tels doubles comptes sont possibles, et finalement 2,59 % en incluant tous les cas possibles de doubles comptes. Sous l'hypothèse que, dix fois sur onze, un seul des logements des personnes en double compte est recensé, notre estimation finale est de 2,47 %. Ramené à l'ensemble de la population (et non plus l'échantillon EDP), cela conduit à une estimation finale de 1 570 000 personnes, pour une population métropolitaine de 63 650 000 au 1^{er} janvier 2013. Les variations aléatoires autour de cette estimation sont du même ordre que précédemment (on observe 7969 bulletins doubles en t+1 de 2011 à 2015), ce qui conduit à un intervalle de confiance final autour de l'estimation de 2,47 % de l'ordre de [2,39 ; 2,55]. Les doubles comptes sont à peu près deux fois plus probables à partir des bulletins doubles dans deux vagues successives, ce qui confirme notre hypothèse selon laquelle les doubles comptes sont plus souvent visibles et donc évités quand les deux logements sont interrogés la même année. (Tableau 3).

Tableau 3 : Comparaison en t;t et t;t+1 des individus déclarant ne pas avoir déménagé (2011-2015)

	Effectifs pondérés		Proportions (en %)	
	t ; t	t ; t+1	t ; t	t ; t+1
a. Individus EDP recensés (pondérés)	13 804 821	13 804 821	100,00	100,00
b. Individus recensés deux fois	174 896	750 501	1,27	5,44
c. Même adresse ou même petite commune	7 615		0,06	
d. Individus déclarant ne pas avoir déménagé* (hors même petite commune ou même adresse)	105 462	221 033	0,76	1,60
e. Estimation d'individus en doubles comptes (hors même petite commune ou même adresse)	167 281	350 597	1,21	2,54
f. Estimation globale	174 896	358 212	1,27	2,59
g. Estimation finale	341 547		2,47	

Champ : t ; t : bulletins doubles des individus nés un jour EDP. t ; t+1 : BI des individus recensés lors de deux vagues successives. EAR 2011-2015.

Source : Base étude 2015 de l'EDP.

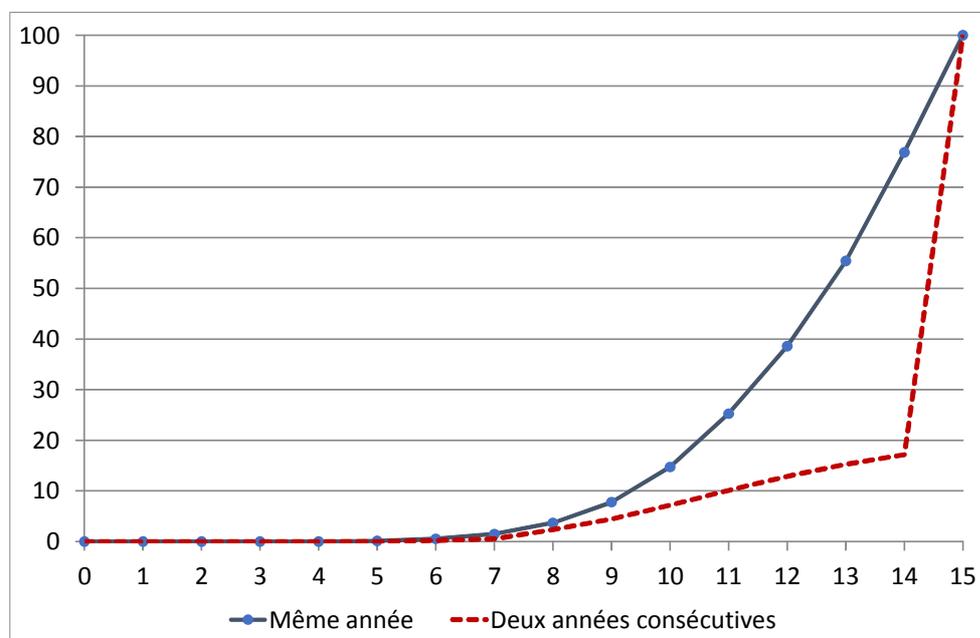
Note : * : Pour t;t, sont également inclus les individus déclarant avoir déménagé dans un des deux logements recensés (on divise alors par deux leur poids pour avoir des champs comparables). Construction du poids pour t ; t+1 : produit du poids en t et en t+1, le tout divisé par 2, car il y a deux manières d'être présents en t et en t+1 si on a deux logements. La comparaison se fait sur le sous-champ des habitants qui n'ont pas déménagé ou qui n'ont déménagé que d'un logement depuis un an. Pour faciliter la cohérence des totaux on estime t;t avec t de 2011 à 2015, et t;t+1 avec t+1 de 2011 à 2015. Méthode d'estimation :

En t;t on observe tous les cas. En t;t+1 on observe a, b et d. On en déduit $e(t;t+1) = d(t;t+1) * e(t;t) / d(t;t)$ puis $f(t;t+1) = e(t;t+1) + c(t;t)$. Finalement $g = [f(t;t) + 10f(t;t+1)] / 11$, estimation valable avec $p = 1/6$ pour les individus en double compte (voir tableau 4, poids = $1/p^2 = 36$ pour les individus en bulletin double).

Nous avons reproduit le même score d'appariement que pour les doubles comptes présents dans une seule vague de recensement. Les individus repérés dans deux vagues successives de recensement, déclarant ne pas avoir changé de logement (N = 7 969) partagent un ensemble de caractéristiques

communes beaucoup plus fréquentes que pour les doubles comptes présents dans une seule année : dans 83 % des cas les 15 variables sont identiques dans les deux logements (Figure 4). On peut supposer que les personnes recensées deux années de suite remplissent plus souvent elles-mêmes les deux bulletins que celles pour lesquelles deux bulletins sont remplis la même année. Les erreurs d'appariement n'ont ainsi qu'un faible impact sur la surestimation des individus comptés deux fois.

Figure 4 : Score d'appariement pour les individus recensés lors de deux vagues successives d'EAR : fréquence cumulée de réponses identiques (en %)



Champ : BI des individus nés un jour EDP déclarant ne pas avoir déménagé entre deux années successives, recensés en t et en $t+1$. EAR 2010-2015. $N=7969$.

Source : Base étude 2015 de l'EDP.

Lecture : parmi l'ensemble des individus déclarant ne pas avoir déménagé entre deux vagues de recensement, 17% ont 14 informations identiques ou moins entre leurs deux bulletins de recensement remplis.

Nos premières estimations peuvent donc dans ce cas être revues à la hausse, puisque les erreurs d'appariement sont moins nombreuses et les probabilités d'être en bulletin double plus élevées quand on considère les personnes recensées en t et en $t+1$. Il en résulte que la proportion de personnes en double compte est de l'ordre de 2,5%. L'intervalle de confiance est fourni à titre indicatif, les variations aléatoires (0,1% en plus ou en moins) étant faibles devant les incertitudes dues à la méthode de mesure. Les incertitudes principales proviennent des hypothèses faites sur la qualité des réponses à la question sur l'ancienneté dans le logement et la capacité des personnes à « éviter » un double compte si leurs deux logements sont interrogés deux années de suite.

L'estimation à partir des bulletins doubles la même année est clairement une estimation biaisée, minimale, puisque dans ce cas le double compte est visible pour les personnes dont les deux logements sont recensés au même moment. L'estimation fondée sur deux années consécutives est fragile : si les habitants ont tendance à surestimer leur ancienneté dans le logement, des déménagements peuvent être considérés à tort comme des bulletins doubles. À l'inverse, si certains doubles comptes sont également évités en $t+1$, plus que dans les autres cas, notre estimation devrait en être augmentée. S'il faut proposer un intervalle de crédibilité qui contienne presque sûrement la vraie proportion de doubles compte, il pourrait être de [1,5% ; 3,0%].

4.5. Profil par âge d'individus en double compte

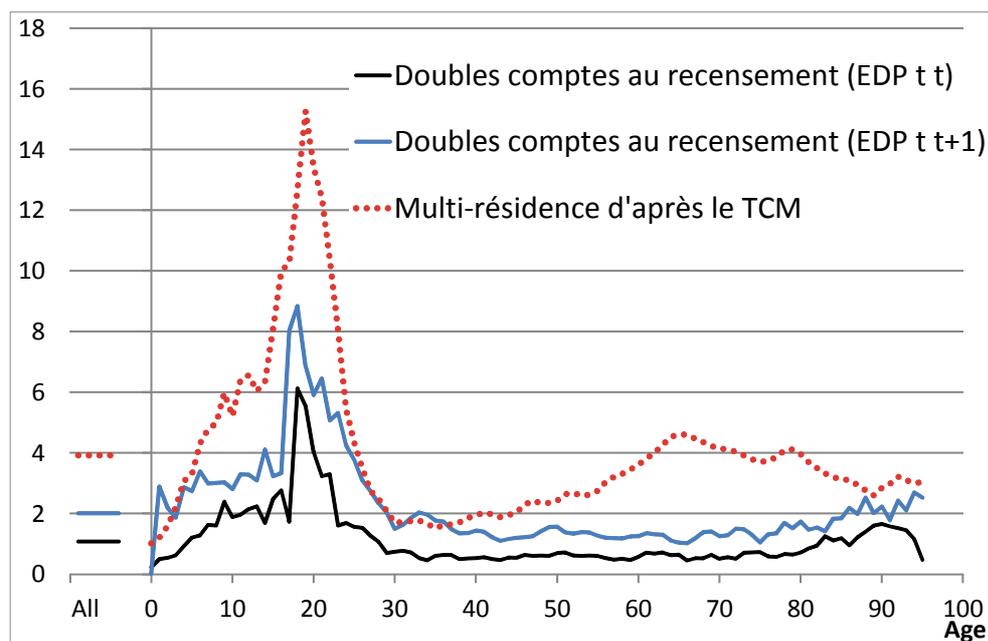
Le profil par âge des individus comptés deux fois au cours d'une seule vague de recensement peut être mis en parallèle au profil de ceux qui sont en situation de multirésidence, bien que la multirésidence ne soit pas systématiquement synonyme de double compte.

Depuis de nombreuses années, la plupart des enquêtes de l'Insee auprès des ménages incluent un ensemble de questions communes, regroupées dans le « Tronc commun des ménages » (TCM). Les fichiers issus de ces différentes enquêtes ont été empilés et permettent d'évaluer précisément la proportion de personnes déclarant « vivre aussi ailleurs » dans un autre logement. Ces personnes sont censées être incluses dans tous les logements où elles résident habituellement, et il est nécessaire de corriger les pondérations de celles dont les deux logements sont éligibles pour obtenir des estimations non biaisées de la proportion de personnes ayant plusieurs résidences dans lesquelles elles sont susceptibles de participer à une enquête (Toulemon et Denoyelle, 2012). Des questions similaires sur la multirésidence ont été incluses dans l'enquête Famille et Logements, réalisée dans le cadre du recensement de 2011. La compréhension des situations de multi-résidences chez les adultes est beaucoup plus extensive que dans le TCM, et inclut des résidences probablement considérées comme secondaires (et donc exclues) au recensement (Imbert et al. 2014), mais la proportion de personnes répondant que leurs enfants vivent aussi dans un autre logement est très proche de celle déduite des enquêtes auprès des ménages (Toulemon, 2018). Les enfants multi-résidents doivent donc, au recensement comme dans les enquêtes auprès des ménages, voir leur pondération divisée par deux si leurs deux logements sont des logements ordinaires.

La multirésidence est un événement de court terme et est plus fréquente en milieu urbain (dans les grandes communes). Dans les cas de situations de multirésidence, les individus ont plus de chances d'être comptés deux fois lorsqu'ils sont très jeunes enfants, à 18 ans avec ce pic au recensement observé, et aux âges élevés.

Les jeunes enfants concernés par les situations de multirésidence sont principalement ceux en résidence alternée, après la séparation de leurs parents. Le risque d'avoir des parents séparés augmente avec l'âge de l'enfant, expliquant cette montée progressive avec l'âge du nombre d'enfants comptés deux fois (Figure 5). La proportion se stabilise aux alentours de 10 ans, pour ensuite connaître un pic marqué à l'âge de 18 ans. Cet âge se caractérise pour ceux qui poursuivent des études par le début de la vie étudiante, associé à un premier départ du domicile parental. Les règles du recensement pourraient alors être plus floues dans ces cas-là, et le jeune recensé à la fois dans son logement étudiant et chez ses parents. Le pic à 18 ans au recensement se retrouve légèrement décalé par rapport au pic des multirésidents (19 ans). La proportion diminue ensuite assez rapidement jusqu'à 35 ans puis est stable, pour connaître un léger regain aux âges élevés (caractérisé par la présence en résidence collective). Les personnes à l'âge de la retraite qui possèdent une résidence secondaire et vivent dans celles-ci une partie de l'année sont rarement recensées en double compte : comme le recensement a lieu en janvier-février, il y a très peu de chances pour que les seniors soient recensés à la fois dans leur résidence d'hiver et dans leur résidence d'été.

Figure 5 : Profil par âge des individus en situation de multirésidence (SRCV) et en double compte (EDP)



Champ : Individus nés un jour EDP. Données pondérées. Données lissées sur trois âges à partir de 25 ans.

Source : Base étude 2015 de l'EDP, Fichiers empilés des TCM 2006-2010. Pondération des enquêtes corrigées des doubles comptes.

4.6. Situation familiale des individus en doubles comptes

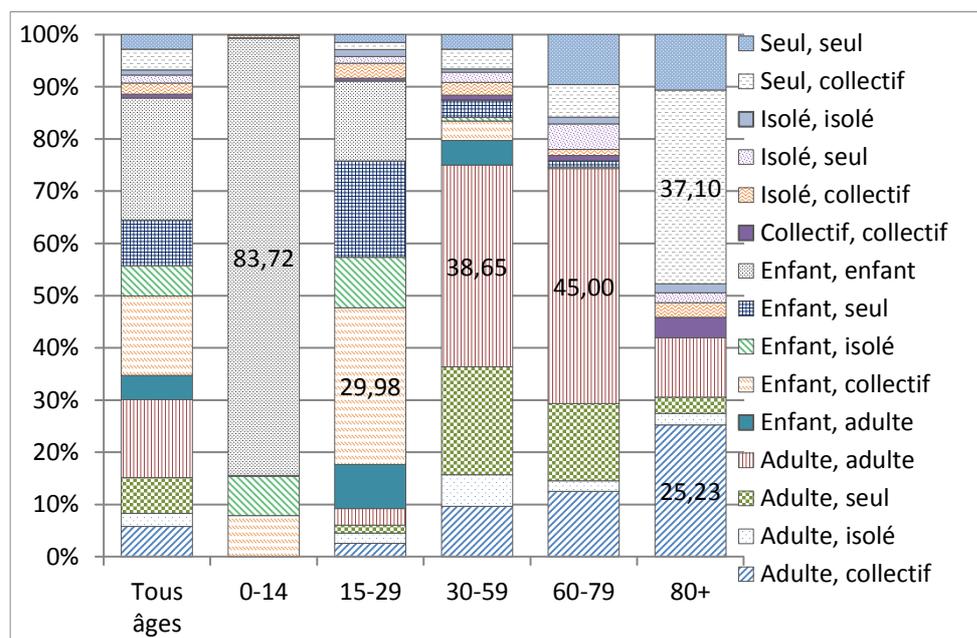
Les doubles comptes posent un problème de comptage des individus, mais les bulletins doubles renseignent sur les situations professionnelles et familiales des habitants qui vivent dans plusieurs résidences habituelles et sont susceptibles d'être recensés dans chacune d'entre elles. Ils permettent également d'étudier les ensembles familiaux à cheval sur plusieurs logements (par exemple de savoir combien de personnes circulent entre les logements). Une analyse complète de ces situations dépasse le cadre de cet article, mais l'observation des positions familiales croisées des personnes dans leurs deux logements, selon le groupe d'âge, offre un premier éclairage sur ces situations, à partir des vagues 2010-2014 de l'EDP (Figure 6).

Dans 56 % des cas, la personne est enfant dans au moins un des deux logements. Cela est principalement dû au fait que les doubles comptes sont élevés aux jeunes âges et lorsque le jeune adulte commence à prendre son autonomie, puisque le pourcentage d'individus de 15-29 ans au moins enfant dans un des deux logements est de 82 %.

Plus précisément, dans 26 % des cas, situation la plus commune, les individus sont enfants dans un des deux ménages et hors famille, hors ménage ou seul dans leur autre logement. Cette configuration est fréquente chez les jeunes adultes de 15-29 ans, où ils étudient dans un lieu et peuvent revenir de temps à autre au domicile parental. Autre configuration possible : les jeunes qui quittent le domicile parental (seul) après avoir trouvé un emploi et sont encore recensés chez leurs parents.

Dans 13 % des cas, l'individu est adulte dans les deux logements, avec ou sans enfant. Majoritaire parmi les adultes de 30-59 et 60-79 ans, cela concerne le plus souvent des couples sans enfant qui sont recensés dans deux logements. Retraités s'ils ont plus de 60 ans, la deuxième résidence recensée pourrait être une résidence secondaire.

Figure 6 : répartition des situations familiales dans les deux logements d'individus en doubles comptes (2010-2014)



Champ : Individus nés un jour EDP, recensés deux fois la même année. Données pondérées. EAR 2010-2014.

Source : Base étude 2015 de l'EDP.

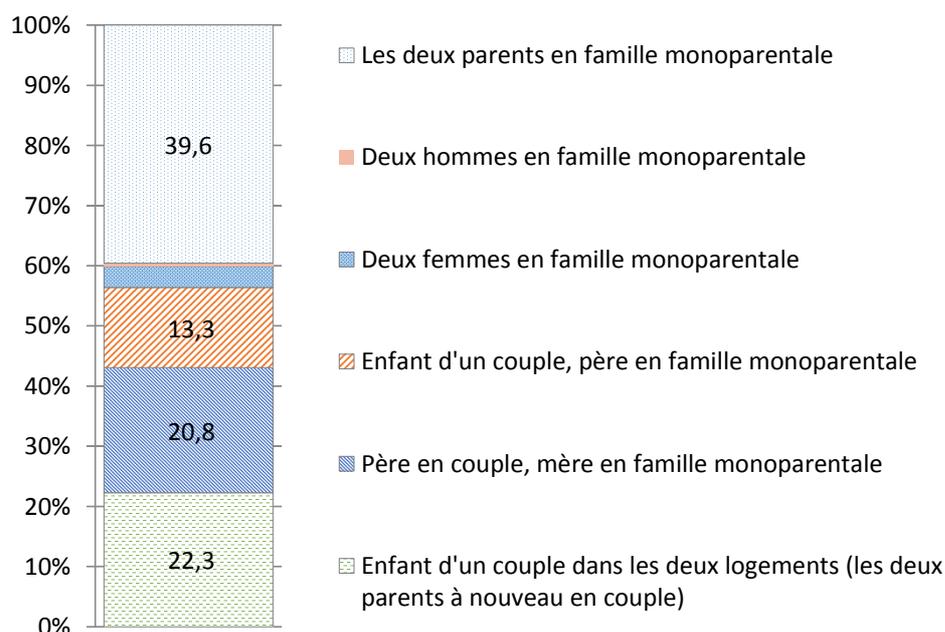
Par ailleurs, certaines personnes sont peut-être recensées à tort seules dans un logement, ce qui pourrait conduire à augmenter artificiellement le nombre de résidences principales, et donc à faire baisser le nombre moyen de personnes par ménage : 19 % des sont seules dans un de leurs logements, et 5 % sont seules dans les deux logements. La manière dont le RIL serait corrigé ou non de ces « fausses résidences principales » l'estimation finale de la population. Pour répondre à cette question il faudrait séparer parmi les doubles comptes les bulletins remplis « à tort » et ceux remplis « à raison ». Les conséquences finales des doubles comptes en termes d'estimation de la population sortent du cadre de cet article.

Les situations des enfants de parents séparés

Plus du quart (26 %) de l'ensemble des personnes recensées deux fois sont enfants d'un couple ou d'une famille monoparentale dans leurs deux logements. Dans 4 cas sur 10 les enfants se partagent entre deux familles monoparentales, dont la personne de référence est un homme dans l'un et une femme dans l'autre, ce qui correspond à des parents séparés qui se partagent la résidence de l'enfant. Dans 1 cas sur 4, les enfants sont « enfants d'un couple » dans les deux logements, probablement après la rupture du couple parental et la remise en couple de chacun des deux parents. Le reste des cas équivaut à des situations asymétriques entre les deux logements, à savoir enfant d'un couple dans l'un et enfant d'une famille monoparentale dans l'autre, un des deux parents s'étant remis en couple.

En se limitant aux enfants mineurs identifiés deux fois au recensement comme « enfants d'une famille » (Figure 7), on constate d'abord la très grande cohérence des résultats : seuls 4 % des cas rendent très improbable le fait que l'enfant ait un de ses parents dans chacun de ses logements : 4,1 % des enfants vivent avec deux parents de même sexe dans les deux logements (deux femmes, probablement leur mère et une grand-mère ou une autre femme (3,5 %) ou deux hommes (0,6 %). Dans tous les autres cas, les parents sont de sexe différents ou vivent en couple : 40 % vivent avec deux parents non en couple, chacun dans son logement ; 34 % vivent avec un parent en couple, l'autre (21 % la mère et 13 % le père) en situation monoparentale ; enfin, 22 % des enfants vivent dans deux familles recomposées, avec un beau-père dans un logement et une belle-mère dans un autre. Le recensement permet d'analyser ainsi les situations précises des enfants dans leurs deux logements familiaux.

Figure 7 : Distribution des situations familiales dans les deux logements d'enfants mineurs recensés comme « enfant d'une famille » dans leurs deux logements (2010-2014)



Champ, source : voir Figure 6.

L'analyse des situations familiales des enfants recensés deux fois renseigne sur leur situation familiale, mais permet aussi de corriger les estimations déduites du recensement. Ainsi, à partir de l'enquête Famille de 2011, on estime le nombre d'enfants vivant dans « deux familles » à 900 000 (Buisson et Lapinte, 2017). Les auteures ont proposé des hypothèses sur le mode de vie de ces enfants dans les deux logements familiaux, en l'absence d'information dans l'enquête. Mais, dans l'enquête comme au recensement, 320 000 enfants sont comptés deux fois. En supposant que ce sont les enfants en résidence partagée qui sont comptés deux fois au recensement, on arrive à une estimation assez différente : 260 000 enfants comptés une seule fois (à raison) et 320 000 enfants en double compte (comptés une fois à raison et une fois à tort, d'après les règles du recensement) sont repérés comme vivant en résidence alternée au recensement et dans l'enquête Famille et Logements (Tableau 7). La répartition détaillée des situations peut être estimée en supprimant les doubles comptes, ou à l'inverse en supposant les bulletins doubles représentatifs des doubles comptes. Cela conduit à des répartitions différentes, mais la révision à la baisse du nombre d'enfants en résidence partagée est robuste.

Tableau 7 : Enfants de parents séparés vivant en résidence partagée

		Insee première 1647			Estimation
Famille 1	Famille 2	Rec+	Mono+	Moyenne	EDP
Recomp	Recomp	290	-	145	128
Recomp	Mono	80	660	370	198
Mono	Mono	530	240	385	254
Total		900	900	900	580

Champ : Enfants de moins de 18 ans, en résidence partagée, vivant avec un seul de leurs parents dans chacun de leurs deux logements.

Source : Insee première 1647 : enquête Famille et logements 2011. EDP : enfants observés en 2010-2014.

Note : à l'enquête Famille on ne connaît pas la situation des enfants dans leur autre logement ; dans l'EDP on a supposé que la répartition des enfants en bulletin double était représentative de l'ensemble des enfants en résidence alternée. Une autre méthode, supprimant les enfants en double compte, fournit des

une estimation plus basse d'enfants dont les deux familles sont recomposées (74 000) et davantage de situations « mixtes (261 000). L'analyse détaillée de ces différences sort du cadre de cet article.

Il en résulte que la proportion d'enfants mineurs vivant avec un seul parent est surestimée au recensement et à l'enquête Famille et logements : 24,8 % d'après l'enquête de 2011 au lieu de 22,4 % si l'on corrige des doubles comptes (Tableau 8).

Tableau 8 : Situation familiale des enfants mineurs en 2011

	En milliers			En %	
	IP1647	EDP		IP1647	EDP
Un parent	3 395	3 075	Un parent	24,8	22,4
Recomposée	945	842	Recomposée	6,9	6,1
Monoparentale	2 450	2 233	Monoparentale	17,9	16,3
Deux parents	10 305	10 625	Deux parents	75,2	77,6
Recomposée	531	547	Recomposée	3,9	4,0
Simple	9 774	10 078	Simple	71,3	73,6
Total	13 700	13 700	Total	100	100

Champ : Enfants de moins de 18 ans.

Source : Insee première 1647 : enquête Famille 2011. EDP : répartition des enfants corrigée des doubles comptes, voir tableau 7.

Note : *Recomposée* : enfants vivant soit avec un parent et un beau-parent (nouveau conjoint du parent), soit avec leurs deux parents et des demi-frères ou demi-sœurs nés d'un seul parent ; *monoparentale* : enfants vivant avec un parent non en couple ; *simple* : enfants vivant avec leurs deux parents et sans demi-frères ou demi-sœurs.

5. Conclusion

L'échantillon démographique permanent, en conservant les différents bulletins remplis par les habitants, permet de repérer des « bulletins doubles » remplis par des personnes dans deux logements différents. On peut en déduire la proportion de personnes en double compte au recensement, au sens où elles résident dans deux logements et sont susceptibles de répondre dans chacun de leurs logements. On peut également considérer les personnes en double compte comme des personnes comptées deux fois au recensement, une fois à raison et une fois à tort.

D'après les bulletins doubles une même année d'enquête de recensement, la proportion d'individus en double compte a connu une hausse modérée entre le recensement de 1990 et le recensement rénové. De 0,7-0,9 % en 1990, la proportion est passée à 1,3 % pour les vagues de recensement de 2004 à 2015. L'utilisation du RIL et l'augmentation attendue de la probabilité d'être confronté à plus de doubles comptes se vérifie ici.

Cette estimation de doubles comptes est cependant soumise à réévaluation. En estimant la part de bulletins doubles sur deux années successives, l'estimation est presque deux fois plus élevée. Avec des vagues d'enquête annuelle, le risque pour une personne ayant deux logements que ses deux logements participent au recensement une année donnée est dix fois plus faible que celui d'être recensés dans un seul de ses logements. La très grande majorité des doubles comptes (dont la proportion est finalement estimée à 2,5 % des habitants, soit 4,9 % des bulletins, dont 2,4 % remplis à raison et 2,4 % remplis à tort) ne donnent donc pas lieu à des bulletins doubles, et sont probablement invisibles pour les habitants qui remplissent le plus souvent un seul bulletin une année donnée.

Les doubles comptes augmentent artificiellement le nombre d'habitants en France, puisque les règles du recensement sont créées de sorte qu'un individu ne soit inclus qu'une seule fois au cours d'une vague. Certains de ces doubles comptes concernent des personnes recensées à tort dans un logement où

elles vivent seules, ce qui conduit à une légère surestimation du nombre de résidences principales : l'impact sur l'estimation de population totale pourrait donc être un peu plus faible si le Répertoire des immeubles localisés ne contient aucun ménage considéré à tort comme une résidence principale. Réaliser un bilan statistique de la qualité de la collecte du recensement nécessiterait d'estimer également les omissions. Les omissions sont plus compliquées à estimer et à corriger dans le contexte du recensement rénové, même si elles devraient être moins nombreuses que dans le recensement ancien. Le recours à d'autres sources de données pour préciser l'ajustement à apporter à l'estimation globale de population pourrait venir compléter nos calculs à l'avenir. Notre estimation des doubles comptes, la première depuis le passage à des enquêtes annuelles, participe cependant à l'évaluation du recensement.

Les estimations internationales conduisent à des estimations variées : 0,35 % en Suisse en 2000, 0,6 % au Royaume-Uni en 2010, 0,9 % en Australie en 2001 de la), 1,5 % au Canada en 2006 (1,5 %), 2,8 % aux États-Unis en 2010 (Census Bureau, 2012; ONS, 2012). Le recensement en France dans sa nouvelle forme a probablement conduit à une hausse des doubles comptes, mais le niveau actuel n'a rien d'exceptionnel.

L'Échantillon Démographique Permanent s'avère donc être un dispositif riche et unique pour étudier les individus en double compte dans le recensement. L'utilisation de sources de données complémentaires au sein de l'EDP, telle que les données sociofiscales, alimentera les réflexions sur les caractéristiques sociodémographiques des individus en double compte ou omis dans le recensement, et est une piste prometteuse à l'avenir.

6. Bibliographie

Abbott O. (2009). 2011 UK Census Coverage Assessment and Adjustment Methodology. *Population trends*, 137(1), 25-32.

Buisson G., Lapinte. (2017). Vivre dans plusieurs configurations familiales ». *Insee première* n°1647, 4 p.

Census Bureau (2012). Summary of Estimates of Coverage for Persons in the United States. *Census Coverage Measurement Estimation Report*. Washington: U.S. Census Bureau.

Coast E., Fanghanel A., Lelièvre E., Randall S. (2016). Counting the Population or Describing Society? A Comparison of English and Welsh and French Censuses. *European Journal of Population*, 32, 165-188. doi: 10.1007/s10680-015-9372-y

Coeffic N. (1993). L'enquête post-censitaire de 1990. Une mesure de l'exhaustivité du recensement. *Population*, 48(6), 1655-1681.

Desplanques, G. (2008). Avantages et incertitudes des enquêtes annuelles de recensement en France. *Population*, 63(3), 477-501.

Faivre S. (2018). Diffusion de deux notes rédigées par Pascal Ardilly relatives à l'Échantillon démographique permanent, note interne Insee n° 2018_4373_DG75-L110 du 31/05/2018.

Insee (2018). Base étude 2016 de l'Échantillon Démographique Permanent. Manuel de l'utilisateur. 476 p. <https://utildep.site.ined.fr/fr/variables/variables-de-l-edp/>

Imbert C., Deschamps G., Lelièvre E., Bonvalet C. (2014). Vivre dans deux logements : surtout avant et après la vie active. *Population et Sociétés*, n° 507, 4 p.

Insee (2012). Les pondérations. *Recensement de la population*. 14 p. <https://www.insee.fr/fr/statistiques/fichier/2383177/fiche-ponderation.pdf>

Jugnot S. (2014). La constitution de l'échantillon démographique permanent de 1968 à 2012. 83p. Insee, Documents de travail N° F1406.

Mambetov D. (2014). Etude des BI multiples associés à une même personne dans les enquêtes annuelles de recensement, à partir de l'EDP (base étude 2011). Insee, note interne, n° 1253/DG75-F170/D M, 9 p.

- ONS (2012). Overcount estimation and adjustment. *2011 Census : Methods and Quality Report*.
- Pan Ké Shon J.-L. (2007). Le recensement rénové français et l'étude des mobilités. *Population*, 62(1), 123-141.
- Pla A. (2009). Bilan démographique 2008: plus d'enfants, de plus en plus tard. *INSEE première*, 1220, 4 p.
- Renaud A. (2007). Estimation of the coverage of the 2000 census of population in Switzerland: Methods and results. *Survey Methodology*, 33(2), 199-210.
- Toulemon L.(2018). Combien de personnes ont plusieurs résidences habituelles en France ?, Ch. 1 inImbert C., Lelièvre E., Lessault D., (eds.), *La famille à distance. Mobilités, territoires et liens familiaux*. Paris: Ined, p. 27-49.
- Toulemon L., Denoyelle T. (2012). La définition des ménages dans les enquêtes françaises: comment tenir compte des multi-résidences ? *présentation aux Xles Journées de méthodologie statistique*, Cité internationale universitaire de Paris, session 26, 16 p. http://jms-insee.fr/jms2012s26_1
- Toulemon L., Pennec S.(2009).Two-home adults and children: long-lasting or transitory family situation? The case of France. XXVI IUSSP International Population Conference, Marrakech, Séance 501, Unions, families and households, 39 p. , <http://iussp2009.princeton.edu/abstracts/93441>
- Toulemon, L., Pennec, S. (2010). Multi-residence in France and Australia: Why count them? What is at stake? Double counting and actual family situations. *Demographic Research*, 23(1), 1-40.