

VALEUR PRÉDICTIVE DES PROLONGEMENTS DE TENDANCE ET DU RECOURS À LA TAXE D'HABITATION DANS LES ESTIMATIONS DE POPULATION LÉGALES

Sébastien HALLÉPÉE¹ (*), Heidi KOUMARIANOS² (*), Hugues LERMECHIN³ (**),
Julien TALBOT⁴ (*)

(*) Insee, Département de la Démographie
(**) Isee

Résumé

Depuis la rénovation du recensement durant les années 2000, obligation est faite à l'Insee de publier chaque année les populations légales de toutes les communes. Pour les communes de moins de 10 000 habitants (petites communes) recensées exhaustivement tous les cinq ans, la difficulté réside dans le rafraîchissement annuel de ces informations. Le problème se pose surtout pour les communes recensées avant la date de référence des estimations. Cette phase d'extrapolation repose sur l'évaluation de deux composantes :

- ✓ l'évolution du nombre de logements de la commune, issue des données de la taxe d'habitation (TH) ;
- ✓ l'évolution du nombre de personnes par logement, observée entre les deux derniers recensements.

Ces deux éléments contribuent à estimer chaque année la population de 40 % des communes de moins de 10 000 habitants, occasionnant, par là-même, l'estimation d'environ 20 % de la population française.

Quelle est la valeur prédictive de la TH dans le cadre de ce double enjeu ? Apporte-t-elle une réelle plus-value à l'estimation des populations ? Les évolutions passées de la taille des ménages se prolongent-elles systématiquement dans le temps ? Peut-on anticiper des situations où cela ne serait pas le cas ? Une de ces deux composantes est-elle prédominante dans les estimations ?

Les études présentées ici sur l'extrapolation de la population des petites communes valident l'innovation du nouveau recensement consistant à recourir à une source administrative, par ailleurs très utilisée par les élus locaux : les populations légales de 7 communes sur 10 sont mieux estimées avec la TH, qu'elles ne le seraient par prolongement linéaire des derniers recensements. Elles confirment par ailleurs l'intérêt d'intégrer un paramètre d'évolution de la taille des ménages dans cette phase.

Néanmoins, ces études ont également mis en lumière les limites et faiblesses de la méthode actuelle. Des ajustements ont déjà été mis en place, pour traiter de manière adéquate certains cas particuliers. Mais ces fragilités peuvent aussi se retrouver plus structurellement dans un ensemble de communes partageant le même profil (très petites communes par exemple) dont une correction au cas par cas n'est pas envisageable.

L'utilisation d'estimateurs alternatifs est en cours d'étude, afin d'apporter à terme des améliorations dans la production de chiffres de population.

¹ sebastien.hallepee @ insee.fr

² heidi.koumarianos @ insee.fr

³ hugues.lermechin @ isee.nc

⁴ julien.talbot @ insee.fr

Abstract

Since 2004 the French census has been based on an annual collection of information successively concerning all municipal territories over a five-year period. Each year, an exhaustive census is conducted of one fifth of the municipalities of fewer than 10,000 inhabitants and a sample survey is conducted in all municipalities of 10,000 inhabitants or more, based on a sample of addresses representing 8% of their dwellings.

For these small municipalities, population data has to be estimated during four years, between two exhaustive census : the method for calculating the weighting depends on the collection date and the population category, but all the municipalities are dealt with in a strictly comparable manner over the five-year period.

For municipalities listed before the median year, the household data is extrapolated by continuing the observed trend and making use of the number of dwellings provided by the local taxes (*Taxe d'habitation*). This innovative choice is validated as that the population of 7 out of 10 municipalities is better estimated using information from local taxes than with a simple trend extension based only on the last census. Considering a parameter which reflects changes in household size appears also relevant.

This paper explains the method used to estimate population in municipalities of fewer than 10,000 inhabitants, and established its strenghts and limits, especially for smaller ones.

Mots-clés

Recensement de la population, Source administrative, Populations légales, Extrapolation, Taxe d'habitation.

Sommaire

Introduction.....	4
1. Principe de la méthode d'extrapolation	6
1.1. <i>L'utilisation de la TH constitue une innovation mais génère de l'approximation pour le calcul des populations.....</i>	6
1.2. <i>Pour compléter l'extrapolation, on utilise la tendance de la taille des ménages de la commune observée entre les deux dernières collectes de recensement</i>	7
2. Valeur prédictive de la source TH pour extrapoler le nombre de résidences principales d'une commune	8
2.1. <i>Comment juger de la valeur prédictive de la TH ?.....</i>	8
2.2. <i>Une erreur d'estimation du nombre de résidences principales inférieure à 5 % pour 3 communes sur 4 avec la TH</i>	10
2.3. <i>Une légère inclinaison à la surestimation pour la méthode d'extrapolation TH qui augmente avec la taille des communes.....</i>	10
2.4. <i>Les faiblesses de la méthode de prolongement linéaire mettent en lumière l'apport de la TH dans le calcul des populations</i>	12
2.5. <i>L'utilisation de la TH permet de mieux estimer le nombre de résidences principales que la méthode de prolongement linéaire pour près de 8 communes sur 10 et rend compte correctement des ruptures d'évolution de séries de logements</i>	13
2.6. <i>L'approximation engendrée par la TH pour les communes estimées par extrapolation dans le cadre des populations légales est vraisemblablement moindre que celle calculée dans l'étude.....</i>	15
3. Valeur prédictive du prolongement de tendance d'évolution de la taille des ménages d'une commune.....	16
3.1. <i>Distribution et moments des tendances d'évolution de la taille des ménages</i>	17
3.2. <i>Qualité de l'estimation en utilisant une prévision de tendance.....</i>	19
3.3. <i>Erreurs d'estimations ayant comme origine une prévision de tendance inadéquate</i>	21
4. Améliorations de la méthode d'extrapolation	23
4.1. <i>Traitements des cas aberrants</i>	24
4.1.1. <i>Les séries de logements TH sont corrigées pour 1 % des communes en extrapolation</i>	24
4.1.2. <i>La tendance d'évolution de la taille des ménages est corrigée en cas de choc sur la structure du parc de logements.....</i>	25
4.2. <i>Nécessité d'une approche globale pour proposer des améliorations de la méthode d'extrapolation.....</i>	28
4.3. <i>Possibilité d'utilisations alternatives des sources actuelles.....</i>	29
5. Conclusion.....	31

Introduction

En France, le dernier recensement général exhaustif de la population date de 1999. Il se tenait alors irrégulièrement : tous les six à neuf ans depuis la fin de la deuxième guerre mondiale. Pour des motifs budgétaires, le recensement de 1999 eut lieu deux ans après la date initialement prévue. A l'issue de celui-ci, un processus de rénovation du recensement a été enclenché, dans le but d'améliorer la qualité du service rendu au public tout en contenant le coût d'une telle opération. Ce processus a abouti en 2004 avec les premières Enquêtes Annuelles de Recensement (EAR), qui se sont substituées au recensement exhaustif. Un dispositif juridique conséquent définit et encadre les nouvelles orientations. L'Institut national de la statistique et des études économiques (Insee) demeure le principal acteur du recensement. Le rôle des communes s'est accru, puisqu'elles sont désormais tenues de préparer et réaliser les enquêtes de recensement sous le contrôle de l'Insee (Godinot, 2005).

Dans ce recensement rénové, on distingue désormais les grandes communes (10 000 habitants et plus) des petites communes (moins de 10 000 habitants). Les grandes communes sont recensées chaque année, mais seule une partie des adresses est enquêtée (environ 8 % d'entre elles chaque année). Dans les petites communes, le recensement exhaustif perdure et se tient tous les cinq ans. Il existe cinq groupes équilibrés de petites communes, qu'on appelle « Groupes de Rotation » (GR) et qu'on recense à tour de rôle année après année. Par exemple, en 2004, les communes du 1^{er} groupe de rotation ont été recensées. En 2005, ce fut au tour du 2^{ème} groupe de rotation et ainsi de suite... A nouveau, en 2009, on a recensé les communes du 1^{er} groupe de rotation, et, en 2010, celles du 2^{ème} ...

Depuis la rénovation du recensement de la population, l'Insee fournit chaque année les populations légales de toutes les communes

L'obligation faite à l'Insee de renouveler chaque année les populations légales de toutes les communes de France⁵ constitue une amélioration notable du recensement rénové. Rien de tel n'était proposé jusque-là. Cela a nécessité de mettre en œuvre de nouvelles méthodes de calcul de la population. Dans les grandes communes, chaque EAR renouvelle l'information et contribue à actualiser l'estimation de la population au 1^{er} janvier d'une année. L'Insee publie les populations légales de manière décalée : par exemple, au 31 décembre 2014, il livre les populations communales en date du 1^{er} janvier 2012. Dans les petites communes, ce décalage est utilisé à escient (cf. Tableau 1). Pour le groupe de rotation recensé l'année de calcul des populations légales et les deux groupes de rotation sur cinq disposant d'une EAR strictement postérieure à la date de calcul des populations légales, l'information du recensement sera privilégiée avant toute autre pour l'estimation de la population. On dit que ces communes sont estimées par « interpolation » ou par les chiffres de collecte. En revanche, pour les deux derniers groupes de rotation, le recensement le plus récent est strictement antérieur à la date de calcul des populations légales. L'Insee utilise alors une source externe d'informations : il s'agit des données de la Taxe d'Habitation (TH) constituées par la Direction Générale des Finances Publiques (DGFIP), qui permettent d'approcher la dynamique d'évolution du nombre de logements et par extension la dynamique de la population dans la commune. On dit de ces communes que leurs populations sont estimées par « extrapolation » ou « extrapolation TH ». Pour simplifier le propos, on parlera aussi dans le reste de l'étude de « communes en extrapolation » (cf. Encadré 1).

⁵ Exceptions faites des communes de Mayotte, de Wallis-Et-Futuna, de la Nouvelle-Calédonie et de la Polynésie Française, où des recensements généraux exhaustifs sont organisés tous les cinq ans.

Tableau 1 : Extrapolation des petites communes pour les populations légales des deux années suivant l'EAR

Année	2010	2011	2012	2013	2014
GR2	collecte	→	2ème extrapolation		
GR3		collecte →	1ère extrapolation		
GR4			collecte		
GR5			2ème interpolation	← collecte	
GR1			1ère interpolation	←	collecte

Après 11 enquêtes annuelles de recensement et la publication de 7 millésimes de populations légales, on dispose de données permettant d'analyser la qualité des éléments utilisés pour l'extrapolation des populations légales des petites communes, et de proposer des améliorations.

Après avoir rappelé le principe d'extrapolation des populations légales des petites communes, on analysera l'apport de la source externe sur la taxe d'habitation vis-à-vis de l'extrapolation. On étudiera ensuite la tendance d'évolution de la taille des ménages, et son impact sur l'extrapolation.

Suite à l'étude de ces deux paramètres, on présentera quelques cas d'école, pour lesquels on peut identifier a priori une difficulté d'extrapolation et la corriger. Enfin, on suggérera des pistes d'amélioration plus structurelles de la méthode actuellement utilisée.

Encadré 1 : la formule de calcul de la population des ménages des communes en extrapolation

Lors d'une extrapolation, la population des ménages est calculée selon la formule suivante :

$$PopMénage_N^{ExtrapolationTH} = PopMénage_{N-1} \times \underbrace{\frac{NbLog_N^{TH}}{NbLog_{N-1}^{TH}}}_{\text{Coefficient TH}} \times \underbrace{\left(\frac{NbPersLog_{DernièreCollecte}^{EAR}}{NbPersLog_{AvantDernièreCollecte}^{EAR}} \right)^{\frac{1}{5}}}_{\text{Tendance d'évolution de la taille des ménages}}$$

où

- $PopMénage_N^{ExtrapolationTH}$ est la population des ménages calculée par la méthode d'extrapolation au 1^{er} janvier de l'année N,
- $PopMénage_{N-1}$ est la population des ménages calculée au 1^{er} janvier de l'année N-1,
- $NbLog_N^{TH}$ est le nombre de logements dans la TH au 1^{er} janvier de l'année N,
- $NbLog_{N-1}^{TH}$ est le nombre de logements dans la TH au 1^{er} janvier de l'année N-1,
- $NbPersLog_{DernièreCollecte}^{EAR}$ est le nombre de personnes par résidence principale calculé lors de la dernière enquête de recensement,
- $NbPersLog_{AvantDernièreCollecte}^{EAR}$ est le nombre de personnes par résidence principale calculé lors de l'avant-dernière enquête de recensement.

L'intervalle intercensitaire étant de cinq ans, on porte à la puissance 1/5^{ème} le rapport des nombres de personnes par logement des deux derniers recensements afin d'obtenir une tendance d'évolution annualisée. La tendance d'évolution de la taille des ménages peut aussi être appelée coefficient de décohabitation en raison du fait que dans une majorité des cas, ce phénomène explique la tendance à la baisse de la taille des ménages au sein des communes.

Remarque : on suppose de façon implicite que la proportion de résidences principales dans le parc de logements est constante sur la période. Bien que son évolution soit irrégulière pour certaines communes, elle est inerte à un niveau agrégé, si bien que l'évolution du nombre de résidences principales est approchée par celle du nombre de logements.

1. Principe de la méthode d'extrapolation

L'utilisation d'une source externe dans le calcul des populations légales a fait l'objet de nombreuses discussions lors de la rénovation du recensement. Plusieurs pistes, comme les fichiers d'Électricité de France (EDF) et de la Sécurité Sociale ont été explorées. Le choix s'est arrêté sur les données de la TH pour plusieurs raisons :

- intuitivement, les concepts de logements au sens du Recensement de la Population (RP) et au sens fiscal semblent proches,
- l'Insee manie les données de la TH depuis longtemps ; il existe donc un savoir-faire à l'institut dans l'exploitation de cette information,
- depuis longtemps, la législation de la taxe d'habitation n'a pas évolué dans un sens qui remettrait en cause la manière dont les locaux sont enregistrés par les services fiscaux ; en conséquence, cela n'a pas causé d'instabilité dans la structure de la base de données.

1.1. L'utilisation de la TH constitue une innovation mais génère de l'approximation pour le calcul des populations

La mise en place d'une base de sondage d'adresses dans les grandes communes a constitué sans nul doute l'élément méthodologique le plus discuté et le plus mis en avant dans la rénovation du recensement. L'utilisation directe d'une source de données externes (en l'occurrence la TH) dans le calcul des populations légales est apparue beaucoup plus anecdotique, alors qu'elle représente pourtant également une innovation importante. Pour la première fois, des données non produites par l'Insee servent au dénombrement de la population française. De plus, la TH contribue chaque année à l'estimation de la population de près de 40 %⁶ des communes, soit environ 20 %⁷ de la population, puisque les petites communes forment la moitié de la population française. L'enjeu lié à la TH est donc double : il faut approcher correctement la population des petites communes en extrapolation tout en estimant le plus précisément possible l'ensemble de la population habitant dans ces petites communes.

Toutefois, l'utilisation d'une source externe dans le cadre d'un projet, et en particulier d'une source administrative à des fins statistiques, génère nécessairement de l'approximation (Desrosières, 2004). Malgré la proximité des concepts en matière de logements entre la DGFIP et l'Insee, la TH ne fait pas exception à cette règle. Les données de la Taxe d'Habitation d'un millésime renvoient la photographie de l'état du bâti dans les communes au 1^{er} janvier de l'année considérée. Elles sont collectées par la DGFIP dans le but de répondre à des problématiques d'ordre fiscal et budgétaire. En d'autres termes, elles décrivent la réalité du bâti sous ces angles et ne sont pas dédiées au dénombrement de la population française.

De son côté, l'Insee s'attelle avec le recensement de la population à décrire notamment la réalité de l'occupation des logements afin de dénombrer le plus justement la population. Si elles peuvent se confondre localement ou ponctuellement, ces deux réalités ne se superposent pas parfaitement.

Les missions de la DGFIP ne sont pas les mêmes que celles de l'Insee et les bases de données reflètent logiquement cette divergence. Pour la limiter au maximum, l'Insee réalise un traitement sur les données TH.

Le traitement central et primordial appliqué par l'Insee sur les données TH est un filtre des logements visant à s'approcher au mieux du concept de logement utilisé au recensement de la population. Dans la base TH, on retrouve en effet l'ensemble du bâti susceptible d'être assujéti à la TH. Or, ce bâti comprend maisons, appartements, mais aussi garages, box, piscines... S'il est facile d'exclure de la liste des logements les locaux dont la nature diffère d'un local d'habitation, il est plus difficile de raffiner le filtre pour les locaux d'habitation eux-mêmes, car certains appartements et maisons définis ainsi au sens fiscal ne sont pas des logements au sens du RP. Ce raffinement est opéré à l'aide de variables de gestion de la DGFIP qui ne décrivent donc pas la réalité du terrain mais une réalité fiscale. L'utilisation des variables de gestion dans la définition du logement constitue donc un

⁶ Entre 38,8 % et 38,9 % pour les millésimes de populations légales 2006 à 2010.

⁷ Entre 20,1 % et 20,3 % pour les millésimes de populations légales 2006 à 2010.

expédient. Mais de la qualité de cet expédient dépend la valeur prédictive de la TH dans le calcul des populations légales.

Le filtre appliqué reste approximatif : on inclut comme on exclut à tort des locaux parmi les logements. En toute logique, les logements de type communauté doivent être exclus des logements, car la TH sert à l'extrapolation de la population des « ménages ordinaires ». Cependant, une parfaite prise en compte de ces logements s'avère impossible. En conséquence, certaines communautés sont présentes dans le stock de logements TH même si la dérive la plus grave réside dans une prise en compte dissymétrique d'une communauté au cours du temps. Cela peut alors avoir une conséquence non négligeable sur l'évolution du nombre de logements TH calculée par l'Insee. De plus, puisque les services fiscaux locaux sont impliqués dans la gestion de la taxe, des particularismes géographiques existent et entraînent une approximation supplémentaire.

En plus de l'approximation causée par le filtre des logements, deux éléments participent à l'approximation dans la valeur prédictive de la TH, sans qu'il soit possible de les amender. D'une part, les différences de concept sur le caractère d'occupation d'un local obligent à extrapoler les résidences principales sur la dynamique de l'ensemble des logements. En outre, des stratégies d'optimisation fiscale peuvent conduire à des différences déclaratives entre le RP et la TH. D'autre part, le décalage de date entre la déclaration fiscale et la période de recensement peuvent générer des doubles comptes ou au contraire des oublis dans le nombre de résidences principales estimé avec la TH.

1.2. Pour compléter l'extrapolation, on utilise la tendance de la taille des ménages de la commune observée entre les deux dernières collectes de recensement

En effet, l'évolution du nombre de logements à la TH donne une information sur l'évolution du volume des logements, mais masque totalement les aspects démographiques de l'évolution de la population.

Pour extrapoler la population des petites communes, on utilise donc un deuxième paramètre, qui représente la tendance d'évolution de la taille des ménages. Celui-ci est établi, pour chaque commune, à partir de la mesure de l'évolution du nombre moyen de personnes par résidence principale, entre les deux dernières collectes de recensement.

Pour les petites communes, il s'agit donc de collectes exhaustives.

Cette mesure, effectuée sur des données exhaustives et passées, est utilisée pour prolonger la tendance lors de l'extrapolation. On prédit le nombre de personnes par logement en faisant évoluer le nombre de personnes par logement de la dernière collecte grâce à la tendance d'évolution de la taille des ménages.

Le tendance d'évolution de la taille des ménages est la variation annuelle du nombre moyen de personnes par logement entre deux collectes. Ses trois principaux déterminants sont :

- **l'évolution de la structure du parc de logement** de la commune (répartition du nombre de logements selon le nombre de pièces par logement), qui fait référence à un changement du paysage urbain de la commune. Par exemple, une commune pavillonnaire peut se densifier suite à la construction d'un ensemble d'immeubles dans un quartier ;
- **l'évolution du profil des ménages de la commune** (répartition du nombre de ménages selon le nombre de personnes par ménage), qui fait référence à un changement de la structure familiale. Le vieillissement de la population et la croissance des recompositions familiales ont ainsi tendance à réduire la taille des ménages ;
- **l'évolution de la structure d'habitation des ménages de la commune** (répartition du nombre de ménages selon le nombre de pièces occupées par ménage), qui fait référence à un changement de la capacité/contrainte contractuelle d'habitation des ménages. A titre d'illustration, l'augmentation des prix de l'immobilier peut amener les familles cherchant un nouveau logement à opter pour un logement plus petit que celui qu'elles auraient pu s'offrir quelques années plus tôt avec le même budget ;

Ces trois déterminants concernent une vision macro du territoire et de son évolution. D'un point de vue micro, c'est l'évolution du profil de chaque ménage occupant chaque logement (démographie du ménage ou changement de ménage occupant) qui concourt à déterminer l'évolution macro.

Du point de vue relativement agrégé qu'est celui du calcul des populations légales, seuls les trois premiers facteurs (structure du parc de logements, profil des ménages et structure d'habitation des ménages de la commune) sont pertinents.

Pour le calcul des populations légales, l'utilisation de la tendance d'évolution de la taille des ménages, telle qu'elle est conçue, n'est judicieuse que dans le cas où ces déterminants connaissent une évolution relativement stable sur la période d'estimation.

L'utilisation de ce paramètre admet aussi quelques limites :

Le prolongement de tendance est une hypothèse forte, en particulier sur les aspects de modifications du parc de logements. Les chocs ponctuels, ou les tendances structurelles sont prolongées de manière indifférenciée. Les chocs, s'ils sont détectés, ne doivent pas être reproduits lors des extrapolations.

On tentera donc d'identifier des situations particulières (programme immobilier important, changement de statut d'une communauté par exemple) et de proposer une tendance d'évolution alternative, plus pertinente.

Par ailleurs, une évolution est sensible à la qualité des collectes concernées. Si une collecte est de moins bonne qualité, ce sont deux évolutions qui seront impactées (et quatre populations légales).

Enfin, ni la tendance d'évolution de la taille des ménages, ni la TH ne prennent en compte l'évolution du statut d'occupation des logements : de façon implicite, la part de résidences secondaires est supposée constante dans l'extrapolation.

De plus, bien que la population des ménages compose principalement la population municipale, les autres catégories de populations incluses dans la population municipale sont estimées, sans que la TH n'intervienne. C'est le cas de la population des communautés et de la population des Habitations Mobiles terrestres et fluviales et Sans-Abris (HMSA). En conséquence, la TH et le prolongement de tendance de la taille des ménages ne sont pas les seuls éléments qui participent à l'évolution de la population des petites communes en extrapolation.

2. Valeur prédictive de la source TH pour extrapoler le nombre de résidences principales d'une commune

La première composante participant à l'estimation de la population des ménages des petites communes estimées alors que le dernier recensement est antérieur à la date d'estimation (cas d'estimation par extrapolation) est la source TH. Cette source est utilisée pour fournir une projection du nombre de résidences principales des communes en phase d'extrapolation. Il paraît légitime d'évaluer l'apport de la TH à l'extrapolation de la population des petites communes. En particulier, la fraîcheur de l'information est-elle réellement un apport ? Compense-t-elle les éventuelles erreurs engendrées par l'approximation nécessaire en termes de concepts et de finalités des sources ?

2.1. Comment juger de la valeur prédictive de la TH ?

Mesurer l'apport de la TH dans le calcul des populations des ménages nécessite de s'affranchir de l'approximation causée par la tendance d'évolution de la taille des ménages. Pour expliquer simplement la méthode d'extrapolation, le coefficient TH sert à estimer un nombre de résidences principales dans la commune tandis que la tendance d'évolution de la taille des ménages permet d'estimer un nombre de personnes par résidence principale, la combinaison des deux coefficients renvoyant une estimation de population. La TH doit donc être jugée sur sa qualité à prédire le nombre de résidences principales.

L'examen de la valeur prédictive de la TH sera fait à travers le nombre de résidences principales estimé. Or, il faut pouvoir confronter cette estimation à un point de référence, soit un recensement. Cela revient à comparer, une année de recensement donnée, les résultats du recensement aux résultats d'une extrapolation TH. La valeur prédictive de la TH est donc appréhendée sur un intervalle de cinq années, soit le nombre d'années d'extrapolation nécessaires à une telle comparaison. Pour le 1^{er} groupe de rotation de petites communes recensées en 2004, on confronte en pratique les informations collectées en 2009 à une estimation portant sur l'extrapolation du point de recensement 2004 par l'évolution du nombre de logements observée dans la TH entre 2004 et 2009 (cf. Tableau 2). En pratique, c'est comme si l'on confrontait une situation réelle à une situation fictive, où on imagine que le recensement n'avait pu avoir lieu.

Tableau 2 : Présentation de la méthode pour juger de la valeur prédictive de la TH dans le 1^{er} groupe de rotation

Année	Situation réelle (point de référence)		Scénario fictif (estimation)	
	État de recensement	Mode de calcul des populations	État de recensement	Mode de calcul des populations
2004	Recensement	Absence de calcul ⁸	Recensement	Recensement
2005		Absence de calcul ⁸		Extrapolation
2006		Extrapolation		Extrapolation
2007		Interpolation ⁹		Extrapolation
2008		Interpolation ⁹		Extrapolation
2009	Recensement	Recensement	Pas de recensement	Extrapolation

Lecture : dans la réalité, les communes du 1^{er} groupe de rotation ont été recensées en 2004 et 2009 ; leurs populations ont été estimées par extrapolation en 2006, par interpolation en 2007 et 2008 et avec le recensement en 2009. Dans un scénario fictif auquel est adossée la méthode d'extrapolation TH, on suppose que le recensement n'a pas eu lieu en 2009 ; entre 2005 et 2009, les populations des communes du 1^{er} groupe de rotation seraient donc estimées par extrapolation.

L'étude porte sur les petites communes de France métropolitaine, soit les communes collectées exhaustivement à deux occasions entre 2004 et 2013, entre 2004 et 2008 puis entre 2009 et 2013. Sont considérées comme « petites » les communes dont la population municipale était inférieure à 10 000 habitants lors du recensement de 1999 et que l'Insee continue de recenser exhaustivement. Cette définition est retenue car le mode de recensement de ces communes a été celui d'une petite commune durant la période analysée, quelle que soit l'évolution de leur population après 1999. Il est strictement nécessaire, que ces communes n'aient subi aucune modification géographique durant la période d'analyse. On exclut donc les communes disparues, créées ou modifiées ainsi que les communes absentes des données de la TH. Les communes des DOM n'ont pas été retenues dans l'analyse, car, la Réunion mise à part, la TH n'y est pas utilisée pour l'extrapolation de la population.

Il sera également procédé à la confrontation de la méthode d'extrapolation TH avec une méthode d'extrapolation alternative de prolongement linéaire des résultats des deux derniers recensements exhaustifs connus que l'on nommera méthode « PL ».

Ces différentes méthodes s'appuient sur un schéma identique à celui utilisé pour juger de la valeur prédictive de la méthode TH usuelle. On imagine un scénario fictif où le recensement n'a pas eu lieu et où il est nécessaire d'extrapoler les populations cinq années durant pour rafraîchir l'information.

La méthode de prolongement linéaire « PL » est utilisée pour extrapoler les populations des communes où la méthode TH ne peut être mise en œuvre. Celle-ci ne peut l'être dans les communes de Guadeloupe, Guyane et Martinique ainsi que dans certaines communes de France métropolitaine sans base fiscale locale (et donc sans TH). Il semble donc opportun de comparer cette méthode à la méthode TH usuelle pour vérifier le bien fondé de l'utilisation d'une source externe de données.

La taille de la commune est une variable structurante du nouveau recensement et influe sur la valeur prédictive de la TH. Elle illustre les différentes analyses. Les communes seront regroupées en quatre

⁸ Les premières populations légales issues du recensement rénové concernaient les populations au 1^{er} janvier 2006. L'Insee n'a pas publié de populations légales pour les années 2004 et 2005.

⁹ La méthode d'interpolation consiste à positionner l'estimation sur la droite qui relie le dernier chiffre de population publié au point de recensement

classes, selon leur population en résidences principales : moins de 250 habitants, entre 250 et 499 habitants, entre 500 et 999 habitants, 1 000 habitants et plus.

On confrontera les différentes méthodes d'estimation dans le cadre des deux enjeux précédemment énoncés : les estimations de population dans les groupes de petites communes extrapolées et l'estimation totale de la population extrapolée. La TH servant à estimer un nombre de résidences principales, on convertit respectivement ces deux enjeux en estimations de résidences principales dans les groupes de petites communes extrapolées et estimation totale des résidences principales.

2.2. Une erreur d'estimation du nombre de résidences principales inférieure à 5 % pour 3 communes sur 4 avec la TH

La proportion de communes, dont l'estimation par la méthode TH usuelle s'écarte peu du point de recensement, est forte. Pour près de la moitié d'entre elles, l'écart relatif n'excède pas 2 % en valeur absolue. Pour 3 communes sur 4, il n'excède pas 5 % et pour 9 communes sur 10, il est inférieur à 10 %. Les différences sont minimales entre groupes de rotation. Les distributions de l'écart relatif entre le nombre de résidences principales recensées et le nombre de résidences principales estimé par la méthode TH sont presque confondues d'un groupe de rotation à l'autre. Pour chacune des courbes, il y a un pic de distribution proche de 0 (soit l'écart relatif nul). La méthode d'extrapolation TH conduit à surestimer le nombre de résidences principales des quatre groupes de rotation étudiés. La surestimation atteint 4 résidences principales en moyenne, hormis pour le 1^{er} groupe de rotation où elle ne vaut que 2 résidences (cf. Tableau 3). Pour plus de la moitié des communes, l'écart absolu de résidences principales est compris entre -5 et 6 unités. C'est un intervalle d'autant plus faible, que la comparaison intervient au bout de cinq années d'extrapolation. L'écart médian est proche de 0.

Tableau 3 : Description des écarts relatifs et absolus entre le nombre de résidences principales recensées et le nombre de résidences principales estimé par la méthode usuelle TH

	Écart relatif (en % des logements recensés)				Écart absolu (en nombre)			
	Moyenne	1 ^{er} quartile	Médiane	3 ^{ème} quartile	Moyenne	1 ^{er} quartile	Médiane	3 ^{ème} quartile
GR1	-0,2	-3,0	-0,1	2,6	2,2	-4,6	-0,2	4,7
GR2	0,4	-2,6	0,3	3,1	4,4	-3,8	0,5	5,9
GR3	0,6	-2,2	0,5	3,1	4,2	-3,3	0,8	6,0
GR4	0,7	-2,2	0,5	3,2	3,6	-3,2	0,9	6,0
GR5	0,8	-2,0	0,6	3,2	3,2	-3,1	1,0	6,0
Ensemble	0,4	-2,4	0,4	3,0	3,5	-3,6	0,6	5,7

Lecture : dans le 4^{ème} groupe de rotation, l'écart relatif moyen atteint 0,7 % ; l'écart absolu moyen y est de 3,6 résidences principales. La médiane de l'écart absolu est à 0,9 résidence principale, ce qui signifie que 50 % des communes ont un écart inférieur à ce niveau et 50 % des communes un écart supérieur.

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

2.3. Une légère inclinaison à la surestimation pour la méthode d'extrapolation TH qui augmente avec la taille des communes

La TH tend à surestimer le nombre global de résidences principales des communes extrapolées

Quels que soient le groupe de rotation et la plage d'extrapolation (2004-2009 pour le 1^{er} groupe de rotation, 2005-2010 pour le 2^{ème},...), l'utilisation de la TH conduit à surestimer globalement le nombre de résidences principales des communes en extrapolation. La surestimation est cependant faible. Au bout de cinq années d'extrapolation, on aboutit à une surestimation d'au plus 1,2 % du nombre total de résidences principales avec la méthode TH usuelle. L'écart lié à la méthode de prolongement linéaire est compris entre 0,8 et 1,8 % (cf. Tableau 4).

Les méthodes d'extrapolation testées conduisent à une estimation trop importante du nombre de résidences principales. Pour la méthode de prolongement linéaire, mécaniquement, cela tient au fait que l'évolution des logements est moins dynamique sur la dernière période intercensitaire qu'elle ne l'était lors de l'avant-dernière. La surestimation est donc de nature conjoncturelle. En ce qui concerne la méthode reposant sur la TH, l'hypothèse d'une double nature structurelle et conjoncturelle de l'erreur ne peut être écartée.

Tableau 4 : Écart relatif entre le nombre de résidences principales recensées et le nombre de résidences principales estimé par les méthodes TH, PL

	Année de comparaison	Méthode TH	Méthode PL
GR1	2009	0,6 %	0,8 %
GR2	2010	1,2 %	1,2 %
GR3	2011	1,1 %	1,4 %
GR4	2012	0,9 %	1,4 %
GR5	2013	0,8 %	1,8 %
Ensemble	2009 à 2013	0,9 %	1,3 %

Lecture : le nombre de résidences principales estimé par la méthode TH est supérieur de 0,5 % au nombre de résidences principales recensées du 1^{er} groupe de rotation en 2009.

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

Si la TH tend globalement à surestimer le nombre de résidences principales, l'effet se retrouve également au niveau des estimations de chaque commune. Pour autant, cet effet diffère d'un groupe de rotation à l'autre (cf. Graphique A). La proportion de communes surestimées avec la méthode TH usuelle augmente avec les groupes de rotation. Pour le 5^{ème} groupe de rotation, 45 % des communes ont ainsi une estimation du nombre de résidences principales par la méthode TH supérieure au recensement. Dans le 1^{er} groupe de rotation, cette proportion n'atteint que 38 % et la part des communes sous-estimées y est supérieure (à 41 %).

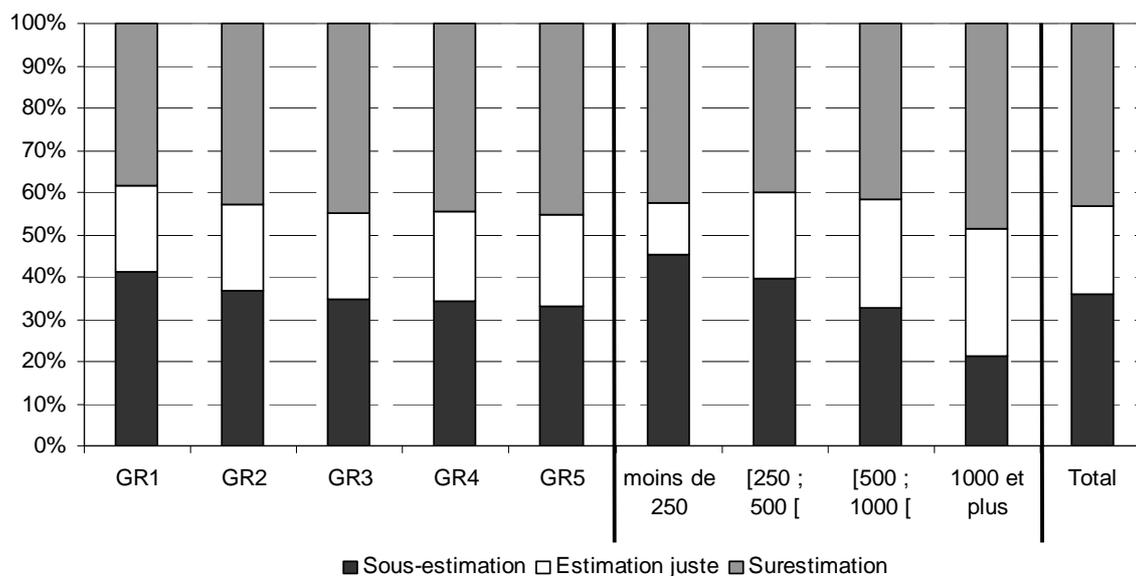
La part des communes justement estimées¹⁰ reste en revanche stable dans le temps. Entre 20 % et 22 % des communes présentent une estimation du nombre de résidences principales par la méthode TH usuelle proche du nombre de résidences principales recensées à 5 % près. L'effet « groupe de rotation » constaté précédemment est confirmé à l'étude de la ventilation des communes selon leur niveau d'estimation par taille de commune. En effet, quelle que soit la taille de commune, la surestimation gagne de plus en plus de communes avec le temps et les groupes de commune. Cette augmentation est parfois conséquente, puisqu'elle atteint 10 points pour les communes de moins de 250 habitants : la proportion de communes surestimées y était de 37 % dans le 1^{er} groupe de rotation, elle y est de 47 % dans le 5^{ème}. Les groupes de rotation étant équilibrés par construction, on peut avancer l'idée que la qualité de l'estimation par la méthode TH dépend de critères conjoncturels : en effet, chaque groupe de rotation est jugé sur une période temporelle différente : 2004-2009 pour le 1^{er} groupe de rotation, 2005-2010 pour le 2^{ème}, 2006-2011 pour le 3^{ème}, 2007-2012 pour le 4^{ème}, 2008-2013 pour le 5^{ème}. Les éléments conjoncturels qui peuvent influencer localement et temporairement sur la qualité de l'estimation peuvent être en lien avec le terrain (dynamique heurtée de constructions de logements, rénovation urbaine, crise économique) ou avec la gestion de l'information (nettoyage de base, utilisation différente d'une variable de gestion qui se répercute sur le comptage des logements par l'Insee). Dans tous les cas, cela se rapporte au mode d'enregistrement de l'information. Par exemple, dans le cas de nouveaux locaux, les services fiscaux peuvent enregistrer un état provisoire la première année avant de le modifier l'année suivante. On peut imaginer aussi que des délais particulièrement serrés une année donnée conduisent les producteurs de l'information TH à des simplifications dans la gestion et la retranscription de l'information. Cela influe sur la qualité de l'information et peut avoir des répercussions sur le dénombrement des logements par l'Insee. L'Insee ne produisant pas les données TH, il semble illusoire d'envisager de contrôler l'aléa relatif à cette composante.

Par ailleurs, on observe que la probabilité de surestimer le nombre de résidences principales par la méthode TH augmente avec la taille de la commune. Cette assertion reste vraie quel que soit le groupe de rotation. La différence entre les proportions de communes sous-estimées dans l'ensemble des communes les plus grandes et celui des plus petites est extrêmement conséquente et reste stable

¹⁰ On considère qu'une commune est bien estimée si le résultat de l'estimation par la TH s'approche du nombre de résidences principales effectivement recensées à plus ou moins 1 %.

d'un groupe de rotation à l'autre. Elle avoisine 20 points. Dans le 3^{ème} groupe de rotation, 44 % des communes de moins de 250 habitants présentent une estimation significativement inférieure au recensement, contre 20 % pour les communes de 1 000 habitants et plus. L'effet « taille de commune » suggère l'existence d'une composante structurelle dans la qualité de l'estimation. Quel que soit le groupe de rotation, le filtre usuel des logements conduit à surestimer plus probablement le nombre de résidences principales des communes les plus grandes.

Graphique A : Répartition des communes selon leur niveau d'estimation par taille de commune et groupe de rotation



Lecture : parmi les communes de 500 à 1 000 habitants, 43 % ont une estimation du nombre de résidences principales par la méthode TH usuelle strictement supérieure à un pourcent de plus que le nombre de résidences principales recensées. Cette proportion atteint respectivement 38 %, 43 %, 45 %, 45 % et 45 % pour les communes des 1^{er}, 2^{ème}, 3^{ème}, 4^{ème} et 5^{ème} groupes de rotation.

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

Il semble donc que deux composantes de nature différente interviennent dans la qualité de l'estimation par la méthode TH usuelle :

- une composante conjoncturelle, en lien avec l'enregistrement de l'information qu'il semble difficile de contrôler.
- une composante structurelle plus marquée que la composante conjoncturelle, que l'on pourrait essayer de contrecarrer avec les outils dont on dispose. Il n'est toutefois pas dit que ces outils soient suffisants pour améliorer la qualité de l'estimation. L'information dont on aurait besoin pour améliorer le filtre et corriger l'effet de structure est peut-être simplement absente de la base TH.

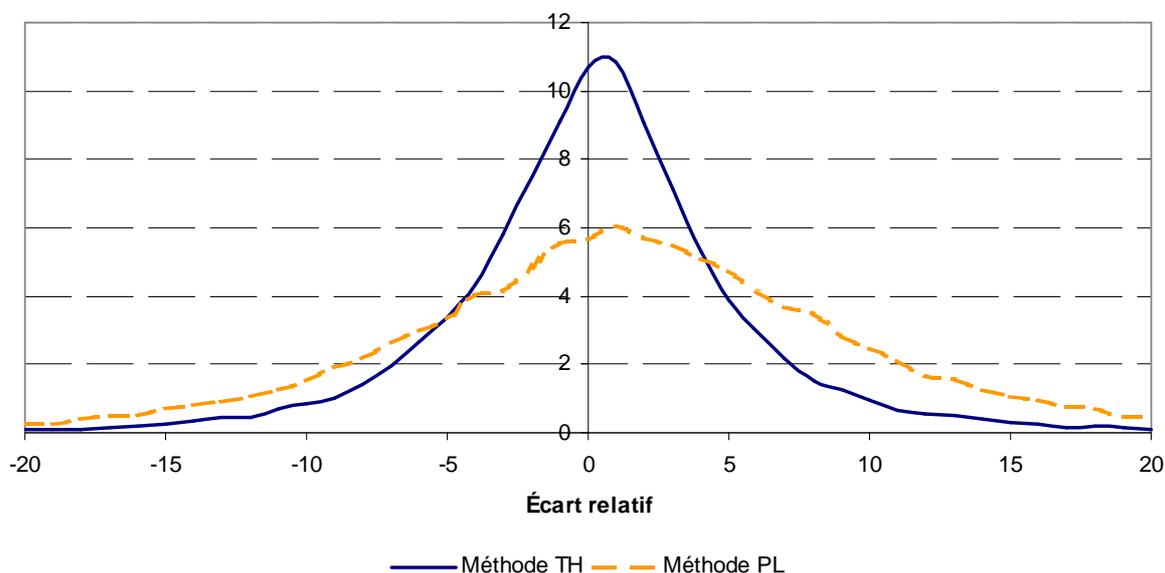
L'identification d'une composante structurelle est intéressante dans l'objectif d'améliorer le codage des logements. Cependant, cet élément ne sera pas expertisé plus avant dans l'étude, dont l'objectif est de décrire et de juger l'apport de la TH et non d'améliorer son usage.

2.4. Les faiblesses de la méthode de prolongement linéaire mettent en lumière l'apport de la TH dans le calcul des populations

En confrontant les différentes méthodes pour l'ensemble des groupes de rotation, on constate que la méthode de prolongement linéaire est de moins bonne qualité que celle utilisant la TH. Elle est notamment bien moins précise, la variabilité des écarts entre le nombre de résidences principales recensées et le nombre de résidences principales estimé par cette méthode étant importante. En

d'autres termes, la méthode TH estime mieux le nombre de résidences principales collectées pour plus de communes que la méthode de prolongement linéaire (cf. Graphique B).

Graphique B : Distribution des écarts relatifs entre le nombre de résidences principales recensées et le nombre de résidences principales estimé par les méthodes TH et PL



Lecture : La proportion de communes dont le nombre de résidences principales estimé avec la méthode TH surestime de 1 % le nombre de résidences principales recensées s'élève à 11 %. Par ailleurs, 2 % des communes possèdent un nombre de résidences principales estimé avec la méthode de prolongement linéaire surestimé de 12 %.

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

La méthode de prolongement linéaire ne permet d'approcher le nombre de résidences principales recensées à 5 % près que pour une commune sur deux. Au contraire, les méthodes recourant à la TH approchent avec la même erreur le nombre de résidences principales recensées pour 3 communes sur 4. Ces résultats sont extrêmement stables d'un groupe de rotation à l'autre.

L'observation des écarts-types confirme l'étendue de la dispersion des estimations pour la méthode de prolongement linéaire et par ricochet sa précision altérée. En moyenne, elle est aussi la méthode qui surestime le plus le nombre de résidences principales (entre 3 et 7 résidences principales). Ces éléments montrent l'imprécision de la méthode de prolongement linéaire relativement aux méthodes employant la TH.

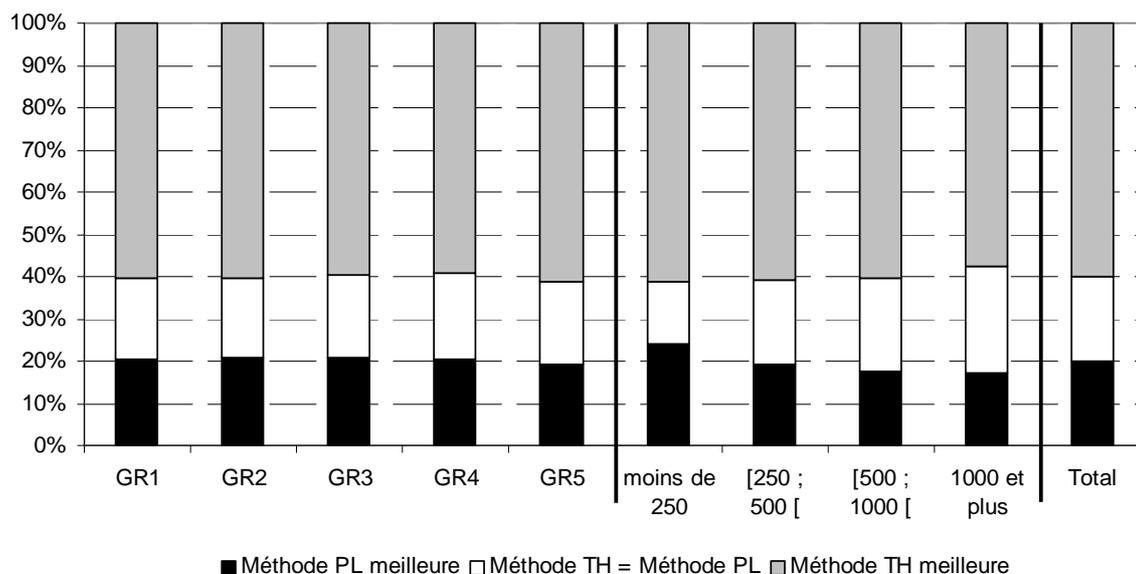
2.5. L'utilisation de la TH permet de mieux estimer le nombre de résidences principales que la méthode de prolongement linéaire pour près de 8 communes sur 10 et rend compte correctement des ruptures d'évolution de séries de logements

Quel que soit le groupe de rotation, la méthode TH usuelle permet d'estimer plus précisément le nombre de résidences principales que la méthode de prolongement linéaire pour plus de 60 % des communes. Dans un peu moins de 20 % des cas, les deux méthodes sont équivalentes. Pour un cinquième des communes en revanche, la méthode de prolongement linéaire est meilleure que la méthode TH (cf. Graphique C).

On considère qu'une méthode est meilleure qu'une autre quand le nombre de résidences principales estimé par la première est plus proche du nombre de résidences recensées que le nombre de résidences principales estimé par la seconde méthode. On compare donc les écarts entre le point de recensement et le point estimé. Deux méthodes sont donc considérées de qualité analogue, si elles renvoient des estimations différentes mais dont les écarts au nombre de résidences principales

recensées sont identiques à 1 % près. Par exemple, pour un nombre de résidences principales effectivement recensées de 1 550, si les méthodes testées renvoient une estimation de 1 540 (-0,6 %) et 1 558 (+0,5 %) résidences principales, elles seront considérées comme de qualité équivalente. En effet, en valeur absolue, l'écart entre les deux méthodes n'est que de 0,1 %. Ce parti pris entraîne par construction une occurrence bien plus élevée des équivalences de méthodes dans les groupes des communes les plus grandes. On observe que plus la commune est peuplée, plus il est fréquent que les méthodes TH et de prolongement linéaire aboutissent à la même estimation près de nombre de résidences principales à 1 % près.

Graphique C : Confrontation des méthodes TH et PL



Lecture : 60,3 % des communes du 1^{er} groupe de rotation sont mieux extrapolées avec la TH qu'elles ne le seraient avec la méthode de prolongement linéaire. Pour 19,4 % d'entre elles, les estimations des deux méthodes sont équivalentes à 1 % près. Enfin, la méthode de prolongement linéaire serait meilleure pour 20,3 % des communes.

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

Le groupe de rotation ne joue pas sur la confrontation des méthodes TH et PL. L'apport de la TH est bénéfique ou neutre dans l'estimation de 80 % des communes quel que soit leur groupe de rotation. Cet axe de confrontation des méthodes fournit un nouvel argument en faveur de l'utilisation de la TH dans le processus de calcul des populations.

Par ailleurs, la méthode TH accentuera son avantage pour les communes les plus grandes. En effet, la TH donne une prévision au moins aussi bonne que la tendance passée pour 75 % des communes de moins de 250 habitants contre 82 % des communes de 1 000 habitants et plus.

La méthode de prolongement linéaire, pourtant utilisée pour le calcul des populations des communes des DOM (à l'exception de Mayotte et de la Réunion) et des communes à défaut de TH (certaines rares communes de France métropolitaine ne disposent pas de TH), accuse certaines limites. Une méthode d'extrapolation du nombre de logements basée sur le prolongement linéaire d'une tendance passée ne peut en effet restituer le heurt constaté pour les communes où existe un retournement de tendance dans la dynamique de logements. Le recours à la TH dans le cadre du calcul des populations légales est notamment motivé par l'hypothèse que ces données permettent de saisir ces retournements. En revanche, on peut se demander si le recours à la TH reste judicieux dans le cas de communes dont les évolutions sont stables.

On convient que l'évolution des logements est stable pour une commune lorsque l'écart entre le nombre de résidences principales estimé par la méthode de prolongement linéaire et le nombre de résidences principales recensées ne dépasse pas 5 %. Cette définition divise les communes en deux groupes d'effectif comparable : il y a une moitié de communes à la dynamique de logements stable (48 %), et une autre moitié à l'évolution heurtée (52 %). Lorsque la dynamique de logements est

stable, cela signifie que la méthode PL prédit correctement tandis qu'elle prédit mal quand la série de logements subit des à-coups. On utilise le même seuil de 5 % pour juger de la proximité de l'estimation du nombre de résidences principales par la méthode TH avec le nombre de résidences principales recensées : si l'écart relatif par la méthode TH usuelle est inférieur ou égal à 5 %, on considère que la méthode TH prédit correctement. Selon ce critère on retrouve la supériorité de la méthode TH qui prédit correctement 73 % des communes.

La confrontation des deux méthodes permet de mesurer la qualité de la TH selon la stabilité de l'évolution des logements (cf. Tableau 5). La méthode TH prédit mieux les communes dont l'évolution de logements est stable, quelle que soit la taille de la commune. Toutefois, à part pour le groupe des communes de moins de 250 habitants, les écarts sont relativement faibles. Dans les communes de moins de 250 habitants, la TH prédit correctement le nombre de résidences principales recensées dans 7 cas sur 10 quand l'évolution est stable et dans un peu moins d'un cas sur deux quand l'évolution est instable. Ainsi, on peut considérer que les perdantes au recours aux données de la TH sont les communes de moins de 250 habitants dont l'évolution est stable : bien estimées par la méthode PL, 3 cas sur 10 ne le sont plus par la méthode TH. En revanche, pour les deux ensembles de communes les plus peuplées à l'évolution heurtée, la méthode TH approche correctement le nombre de résidences principales recensées dans plus de 4 cas sur 5, contre un peu plus de 90 % dans les communes à l'évolution stable.

Tableau 5 : Proportion de communes estimées correctement par la méthode TH usuelle selon la taille de la commune et la valeur prédictive de la méthode PL

	Évolution stable des logements impliquant une bonne estimation PL	Évolution instable des logements impliquant une mauvaise estimation PL
< 250 hab.	69 %	45 %
[250 ; 500 [hab.	86 %	66 %
[500 ; 1000 [hab.	91 %	78 %
>= 1000 hab.	94 %	82 %
Tous	85 %	61 %

Lecture : Dans le groupe des communes de 1 000 habitants et plus, la proportion de communes correctement estimées avec la méthode TH usuelle (à 5 % près) s'élève à 82 % lorsque l'évolution du nombre de résidences principales est heurtée (mauvaise valeur prédictive de la méthode PL) et à 94 % lorsqu'elle est stable (bonne valeur prédictive de la méthode PL).

Champ : petites communes de France métropolitaine estimées par extrapolation TH.

Source : EAR, TH.

En somme, la TH estime de manière plus précise les communes à l'évolution linéaire, mais la dégradation de l'estimation pour les communes à l'évolution heurtée reste très limitée. Ce constat n'a rien de choquant, si l'on considère qu'il est plus facile de mesurer des états inertes ou stables, quel que soit le contexte. Il est même plutôt encourageant compte tenu des écarts observés. Pour les situations plus heurtées, le recours à la TH dans le cadre du calcul des populations légales est précieux. Les retournements de tendance sont souvent correctement appréhendés. La TH s'avère donc être un bon instrument pour les estimations des communes en extrapolation, car elle améliore nettement les prévisions en comparaison de celles de la méthode de prolongement linéaire.

2.6. L'approximation engendrée par la TH pour les communes estimées par extrapolation dans le cadre des populations légales est vraisemblablement moindre que celle calculée dans l'étude

La méthode utilisée dans l'étude juge de la qualité de la TH sur une durée cumulée et extrapolée de cinq ans. Si l'on convient qu'une extrapolation une année donnée se révèle d'autant plus hasardeuse que le point de référence s'en trouve éloigné, le jugement porté dans l'étude exagère le niveau d'erreur relatif à la TH pour les extrapolations à 1 et 2 années qui sont celles qui comptent pour le calcul des populations¹¹. Il est impossible d'estimer le niveau de l'erreur lors de ces années en ne

¹¹ A ce titre, l'exemple récent du recensement allemand est particulièrement intéressant. Le recensement de 2011 s'est tenu plus de 20 ans après les précédents recensements qui dataient d'avant la réunification : en 1987 en RFA et en 1981 en RDA. Dans l'intervalle, l'emploi de registres administratifs a contribué à extrapoler la population. Toutefois, le recensement de 2011 a conduit à une réévaluation à la baisse de 1,5 million d'habitants

recourant pas à de nouvelles hypothèses. On peut en revanche avancer raisonnablement que les tendances constatées (notamment la surestimation des résidences principales ou l'apport bénéfique de la TH dans l'estimation de la population) restent vraies quelle que soit la durée de l'extrapolation et que l'approximation engendrée par la TH doit être moindre en réalité, qu'elle ne l'est dans l'étude.

Enfin, pour les communes en extrapolation, l'Insee mène chaque année un long et minutieux examen des évolutions de nombre de logements. Les Directions Régionales sont sollicitées. On vérifie que les communautés ou des logements déjà recensés ne sont pas pris en compte dans l'évolution du nombre de logements. On corrige les effets de gestion significatifs que l'on juge artificiels et qui déforment le comptage des logements. Pour les communes, dont l'évolution est essentiellement portée par les logements non principaux, on privilégie la série des résidences principales et non celle des logements pour le calcul du coefficient TH. Finalement, on corrige chaque année le coefficient TH d'environ 1 % des communes en extrapolation. Cette expertise menée uniquement sur les communes en extrapolation ne peut être évaluée, bien que l'on prétende assurément améliorer les estimations de résidences principales avec les corrections proposées (on ne corrige en effet que lorsqu'on a la certitude que l'erreur est réelle et significative). Il existe donc une partie du processus de calcul des populations légales non mesurable qui participe à l'amélioration de l'utilisation de la TH, car l'examen individuel des communes conduit probablement à diminuer l'erreur associée à la méthode utilisant la TH.

Ainsi, quelle que soit l'évolution réelle du parc de logements de la commune, l'utilisation de la taxe d'habitation permet une meilleure estimation dans la majorité des cas (en comparaison avec le prolongement linéaire de la tendance observée par le passé). Les risques associés à l'utilisation d'une source externe et aux différences de concepts sont donc non avérés ou faibles, et la fraîcheur de l'information est un apport intéressant.

3. Valeur prédictive du prolongement de tendance d'évolution de la taille des ménages d'une commune

L'apport de la TH dans les estimations du RP a donc été confirmé. Cependant, son utilisation actuelle renseigne sur l'évolution du nombre de résidences principales, pas encore sur celle de la population. Pour passer à une évaluation de la dynamique de population de chaque commune, il faut également estimer l'évolution attendue de la taille des ménages sur la période d'extrapolation (cf. encadré 1). D'un point de vue global, celle-ci a tendance à diminuer, ce qui est illustré en partie par le phénomène de décohabitation. Ne pas tenir compte de ce phénomène amènerait donc à surestimer la population des petites communes dont le calcul est fait par extrapolation. Pour éviter cet écueil, un autre élément est donc utilisé dans l'extrapolation des populations des petites communes, issu des données du recensement. La collecte de ces informations étant maîtrisée par l'Insee, les enjeux de différences de concept ou de champ que l'on a pu évoquer pour les logements et la TH sont absents dans ce cas. Cependant, après quelques années de recensement, on peut s'intéresser à la prise en compte de ce phénomène, à ses limites, et à la détection et correction « d'erreurs » potentielles. L'étude du paramètre mesurant l'évolution de la taille des ménages s'intéresse aux mêmes communes que dans la partie précédente, à savoir l'ensemble des communes de métropole n'ayant pas connu de changement de géographie et dont la population est restée sous les 10 000 habitants depuis 1999.

On s'intéresse dans un premier temps à une description statistique de la tendance d'évolution de la taille des ménages.

Ensuite, pour les communes ayant été déjà interrogées deux fois, on mesurera l'erreur effectuée en utilisant une prévision de tendance (à la place de la tendance réelle, que l'on ne connaît pas lors du calcul), ainsi que ses conséquences sur les estimations de population.

de la population. La surestimation a été essentiellement causée par la prise en compte de migrants non radiés des registres de « Länder » où ils n'habitaient plus (Héran, 2013). Il est plausible que la détérioration de l'estimation ait été progressive dans l'exemple allemand avec l'accumulation des non-radiations. L'erreur relative à l'utilisation de la TH pourrait être de même nature en France : accumulative et dérivant progressivement avec le temps.

On s'intéresse par la suite également aux conclusions obtenues selon la taille des communes. Le découpage en tranche de taille retenu est le suivant (les chiffres sont ceux du 3^{ème} groupe de rotation, celui de la collecte 2006, mais les groupes varient peu d'une EAR à l'autre) :

Tableau 6 : Répartition des communes du troisième groupe de rotation selon leur taille

Tranche de taille	Nombre de communes
Moins de 100 habitants	794
100 à 249 habitants	1 881
250 à 499 habitants	1 599
500 à 999 habitants	1 346
1 000 habitants et plus	1 510
Total	7 130

Lecture : Parmi les 7 130 commune du troisième groupe de rotation, 794 comptaient moins de 100 habitants.

Source : Insee, RP.

Champ : petites communes de France métropolitaine estimées par extrapolation.

Ce découpage a été choisi de façon à avoir des groupes contenant suffisamment de communes pour que les statistiques soient pertinentes : il faut noter que les toutes petites communes (moins de 50 habitants) sont très sensibles à la moindre évolution, le simple jeu des arrondis n'étant pas neutre.

3.1. Distribution et moments des tendances d'évolution de la taille des ménages

Au moment de l'étude, les 5 groupes de rotation avaient été recensés au moins à 3 reprises depuis 1999 (1999 compris), le 1^{er} groupe avait même été recensé 4 fois.

On dispose donc de 11 mesures de la tendance d'évolution de la taille des ménages :

- pour le GR1, on dispose de la tendance calculée entre 1999 et 2004 et utilisée pour les estimations de populations légales 2005, puis entre 2004 et 2009 (pour les estimations 2010 et 2011), et enfin entre 2009 et 2014 (qui servira pour les estimations de population 2015 et 2016) ;
- pour le GR2, on dispose de la tendance calculée entre 1999 et 2005, puis entre 2005 et 2010 ;
- pour le GR3, on dispose de la tendance calculée entre 1999 et 2006, puis entre 2006 et 2011 ;
- pour le GR4, on dispose de la tendance calculée entre 1999 et 2007, puis entre 2007 et 2012 ;
- pour le GR5, on dispose de la tendance calculée entre 1999 et 2008, et de celui calculé entre 2008 et 2013.

On se ramène à une évolution annuelle pour que les évolutions soient comparables. C'est d'ailleurs cette évolution annuelle qui est utilisée lors de l'extrapolation.

Les calculs de tendance d'évolution établis à partir du dernier recensement exhaustif de 1999 ont porté sur des périodes de durée différentes selon le groupe de rotation de la commune. En étudiant la distribution de ces 11 mesures, toutes réalisées sur un peu plus de 7 000 communes, on pourra mesurer si l'allongement de la durée entre deux points de collecte a un effet sur la tendance observée. Pour prolonger l'analyse, on peut également simuler ce qu'auraient été les tendances d'évolution de la taille des ménages calculées entre 1999 et la deuxième EAR, ou même la troisième pour le premier groupe de rotation.

Tableau 7 : Éléments sur la dispersion des tendances d'évolution annuelle de la taille des ménages calculées entre les deux collectes les plus récentes

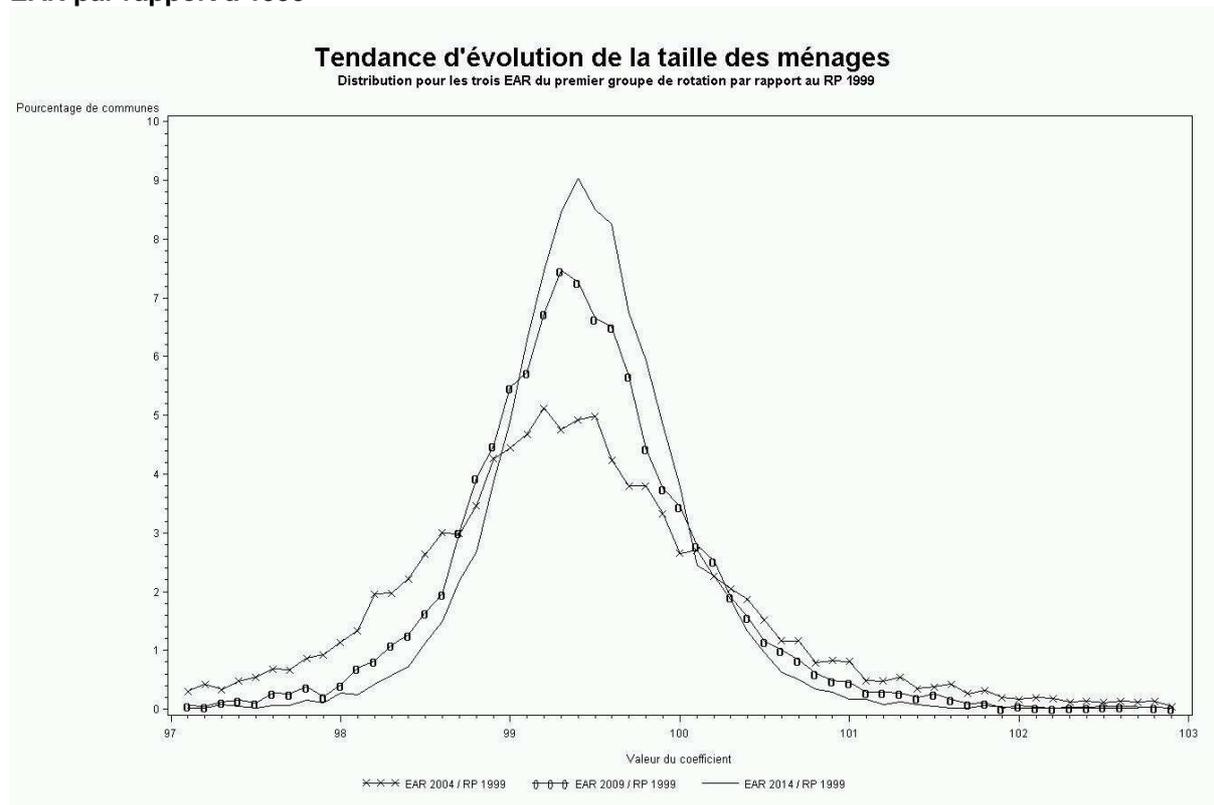
Période de calcul	1 ^{er} décile	Moyenne	Médiane	9 ^{ème} décile	Écart-type
1999->2004	-1,85 %	-0,66 %	-0,69 %	0,57 %	1,15 %
1999->2005	-1,70 %	-0,63 %	-0,65 %	0,49 %	0,99 %
1999->2006	-1,55 %	-0,59 %	-0,63 %	0,43 %	0,90 %
1999->2007	-1,49 %	-0,59 %	-0,61 %	0,34 %	0,82 %
1999->2008	-1,42 %	-0,59 %	-0,61 %	0,28 %	0,77 %
2004->2009	-1,64 %	-0,49 %	-0,53 %	0,72 %	1,07 %
2005->2010	-1,60 %	-0,50 %	-0,54 %	0,66 %	1,07 %
2006->2011	-1,66 %	-0,52 %	-0,54 %	0,65 %	1,04 %
2007->2012	-1,64 %	-0,52 %	-0,53 %	0,63 %	1,03 %
2008->2013	-1,65 %	-0,52 %	-0,52 %	0,60 %	1,05 %
2009->2014	-1,57 %	-0,47 %	-0,48 %	0,64 %	1,02 %

Lecture : Entre 1999 et 2004, la moyenne de l'évolution annuelle de la taille des ménages observée sur les communes du premier groupe de rotation est de -0,66 %. La moitié de ces communes affiche une évolution supérieure à -0,69 %. La dispersion de ce paramètre, évaluée par l'écart-type est de 1,15 % : c'est la période où on mesure le plus de variabilité entre les communes.

Sources : RP 1999, EAR 2004 à 2014.

Champ : petites communes de France métropolitaine estimées par extrapolation.

Graphique D : Distribution de la tendance d'évolution de la taille des ménages pour toutes les EAR par rapport à 1999



Lecture : Entre 1999 et 2014, pour 9 % des communes, la tendance d'évolution annuelle de la taille des ménages est entre -0,5 % et -0,7 % (coefficient de 99,4 %, après arrondi à 0,2 %).

Sources : RP 1999, EAR 2004 à 2014.

Champ : petites communes de France métropolitaine estimées par extrapolation.

A durée intercensitaire constante, les distributions de la tendance d'évolution de la taille des ménages pour chaque EAR sont assez proches (cf. Tableau 7).

On ne constate pas d'évolution temporelle très importante. La tendance moyenne correspond à une diminution de la taille des ménages (d'environ 0,5 % par an). C'est une tendance assez générale et de long terme.

Lorsqu'on s'intéresse à l'influence de l'intervalle intercensitaire sur ce paramètre, on constate que la dispersion est effectivement plus faible lorsque l'écart temporel entre les deux points utilisés grandit. Entre une tendance d'évolution calculée entre deux EAR ou bien entre la deuxième EAR et 1999, on note une réduction de la variance de 50 %. Les intervalles interquartiles et inter-déciles sont réduits, de même que l'étendue (différence entre la valeur la plus élevée et la valeur la plus faible). L'allongement de la durée utilisée pour calculer l'évolution a un effet stabilisateur et réduit la largeur de la distribution (cf. Graphique D).

En effet, au niveau communal, l'allongement de la durée a un effet « lissant » sur l'évolution du nombre de personnes par logement. Ainsi, une commune qui aurait eu une évolution très faible entre 1999 et 2004, puis très forte entre 2004 et 2009, aura une évolution « moyenne » entre 1999 et 2009. C'est le cas assez fréquemment pour les communes les plus petites ayant des évolutions extrêmes (cf. Tableaux 8 et 9).

In fine, l'allongement de la durée « gomme » au moins en partie, les valeurs extrêmes.

3.2. Qualité de l'estimation en utilisant une prévision de tendance

Pour les cinq groupes de rotation, on dispose de deux EAR (et même de trois pour le premier groupe). On peut donc calculer les tendances d'évolution de la taille des ménages entre le RP 1999 et la première EAR, puis la tendance d'évolution entre les deux EAR. La première est utilisée comme prévision de la seconde, on peut donc regarder « l'erreur » commise en prenant (faute de mieux lors des calculs) cette première tendance pour l'extrapolation.

Il est bien évident qu'au moment de l'extrapolation, l'évolution réelle de la taille des ménages n'est pas connue. On cherche ici à savoir si celle qui est utilisée fournit une bonne prévision et si l'on peut identifier les éventuelles causes d'une erreur de prévision importante.

Pour la moitié des communes, l'écart entre les deux évolutions est inférieur à 0,001 (soit moins de 0,1% de la valeur moyenne de ce paramètre). Pour une majorité des communes, le prolongement de la tendance passé fournit donc une estimation qui sera bien confirmée par les résultats de collecte ultérieurs.

Pour 90% des communes l'écart est inférieur à 0,025. A titre de comparaison, l'écart type est de 0,01 et l'intervalle inter-déciles de 0,02.

Pour un peu plus de la moitié des communes (selon les années, 55 à 65 %), ces deux évolutions ont le même signe (et indiquent donc une croissance, ou une décroissance de la taille moyenne des ménages).

Il y a une légère corrélation (significative) entre les deux tendances, le coefficient de corrélation étant de - 0,1.

Ce coefficient varie selon la taille de la commune, entre autre il est positif pour les petites communes de plus de 1 000 habitants.

La bonne qualité prévisionnelle est corrélée à la taille de la commune : plus la commune est petite, plus les événements démographiques sont par nature ponctuels. Si l'évolution de la taille des ménages est le reflet d'un événement ponctuel, il n'est donc pas souhaitable de le reproduire artificiellement par la suite.

La distribution de la tendance d'évolution étant assez concentrée, les écarts entre la tendance passée et celle qui est observée suite à une nouvelle collecte de recensement auront un impact assez limité. Les cas les plus problématiques se concentrent surtout sur les tendances d'évolution qui s'écartent de la moyenne des communes. Ainsi, si on s'intéresse aux valeurs extrêmes de la tendance d'évolution de la taille des ménages, là encore, les petites et grandes communes se comportent différemment. La corrélation négative pour les petites communes provient en partie de ces cas extrêmes : une tendance extrême une année donnée se traduit le plus souvent par un retournement pour la période suivante. En revanche pour les plus grandes des petites communes, les tendances, même extrêmes, se confirment et sont le signe d'un phénomène plus structurel. Néanmoins, mêmes pour les communes de 1 000 habitants et plus, ces retournements de tendance ne sont pas un phénomène marginal.

Tableau 8 : Répartition dans les déciles de la tendance d'évolution de la taille des ménages lors de la deuxième période, en fonction de la tendance de la première période, pour les communes de moins de 100 habitants

		Décile de la tendance d'évolution de la taille des ménages à la deuxième EAR									
		1	2	3	4	5	6	7	8	9	10
Décile de la tendance d'évolution de la taille des ménages à la première EAR	1	7,1 %	8,0 %	7,7 %	8,3 %	8,9 %	8,6 %	10,2 %	10,2 %	15,4 %	15,7 %
	2	6,8 %	9,0 %	9,3 %	11,5 %	6,5 %	11,5 %	12,4 %	9,9 %	9,9 %	13,0 %
	3	6,5 %	7,7 %	9,3 %	7,1 %	12,1 %	11,5 %	10,5 %	10,8 %	10,2 %	14,2 %
	4	7,1 %	7,1 %	8,3 %	10,5 %	13,0 %	10,5 %	7,4 %	11,4 %	11,7 %	13,0 %
	5	5,6 %	11,2 %	10,2 %	9,0 %	8,7 %	12,7 %	13,4 %	10,9 %	9,9 %	8,4 %
	6	11,1 %	11,1 %	9,9 %	9,3 %	11,1 %	11,1 %	10,5 %	8,7 %	10,8 %	6,2 %
	7	11,1 %	10,8 %	7,1 %	12,7 %	10,8 %	8,6 %	7,1 %	11,4 %	9,6 %	10,8 %
	8	14,9 %	9,6 %	14,6 %	10,2 %	11,5 %	11,1 %	8,0 %	9,0 %	6,2 %	5,0 %
	9	14,3 %	9,3 %	10,2 %	11,5 %	8,1 %	9,6 %	10,9 %	12,1 %	9,9 %	4,0 %
	10	16,5 %	16,2 %	12,8 %	10,3 %	9,0 %	8,1 %	6,5 %	5,9 %	5,6 %	9,0 %

Lecture : les pourcentages sont des pourcentages en ligne.

Ainsi les communes pour lesquelles la tendance d'évolution de la taille des ménages 1999-EAR1 appartenait au premier décile se répartissent de la façon suivante : 7,1% de ces communes ont toujours la tendance EAR1-EAR2 dans le premier décile, ... 15,7% sont maintenant dans le dixième décile.

Les cases orange sont supérieures à 12%, celles en vert sont inférieures à 8%.

Source : RP 1999, EAR 2004-2013.

Champ : petites communes de France métropolitaine estimées par extrapolation.

Tableau 9 : Répartition dans les déciles de la tendance d'évolution de la taille des ménages lors de la deuxième période, en fonction de la tendance de la première période, pour les communes de plus de 1000 habitants

		Décile de la tendance d'évolution de la taille des ménages à la deuxième EAR									
		1	2	3	4	5	6	7	8	9	10
Décile de la tendance d'évolution de la taille des ménages à la première EAR	1	16,5 %	14,0 %	10,7 %	8,9 %	9,3 %	7,4 %	7,9 %	7,9 %	8,3 %	8,9 %
	2	12,3 %	12,9 %	11,6 %	9,3 %	13,4 %	8,3 %	8,8 %	7,8 %	7,3 %	8,3 %
	3	11,4 %	10,8 %	10,9 %	11,9 %	8,1 %	10,0 %	9,8 %	10,6 %	7,8 %	8,6 %
	4	11,4 %	9,0 %	10,1 %	9,3 %	10,1 %	10,8 %	11,3 %	10,3 %	11,1 %	6,6 %
	5	7,1 %	10,8 %	10,5 %	10,6 %	9,6 %	11,8 %	12,3 %	9,6 %	8,1 %	9,5 %
	6	9,1 %	7,9 %	11,3 %	12,1 %	9,9 %	10,4 %	10,3 %	9,1 %	10,1 %	9,8 %
	7	10,8 %	9,1 %	7,0 %	11,8 %	9,1 %	12,1 %	8,8 %	9,0 %	10,8 %	11,6 %
	8	8,1 %	7,6 %	10,9 %	9,8 %	11,9 %	9,0 %	10,9 %	11,4 %	10,3 %	10,0 %
	9	8,3 %	8,6 %	9,5 %	8,1 %	9,6 %	9,3 %	9,3 %	12,1 %	12,8 %	12,4 %
	10	9,1 %	8,6 %	7,1 %	7,6 %	8,3 %	10,8 %	10,1 %	11,8 %	13,0 %	13,5 %

De façon plus générale, 9,1 % des communes les plus petites changent d'au moins 8 déciles, alors que c'est le cas pour seulement 5,1 % des plus grandes des communes de moins de 10 000 habitants. 32 % des communes les plus grandes changent au maximum d'un décile, alors que ce n'est le cas que pour 25 % des plus petites.

Dans les plus grandes des petites communes (entre 1 000 et 10 000 habitants), les tendances, surtout les plus extrêmes, semblent donc plus pérennes que dans les petites.

3.3. Erreurs d'estimations ayant comme origine une prévision de tendance inadéquate

On peut également calculer a posteriori l'estimation de la population avec la « vraie » tendance d'évolution de la taille des ménages, et regarder « l'erreur » effectuée en termes de population lors de l'extrapolation.

On peut distinguer deux types d'erreur :

- erreur à évolution TH inchangée : c'est l'erreur que l'on effectue en utilisant une tendance d'évolution de la taille des ménages calculée sur le passé, à la place de la tendance réelle. On ne peut bien sûr mesurer cette erreur qu'a posteriori, lorsqu'on dispose du point de collecte suivant.

Autrement dit cette erreur vaut :

$$erreur_n^{leg} = (nb_log_{n-p}^{EAR} \times evo_nb_log) \times \left(\frac{pop_{n-p}^{EAR}}{nb_log_{n-p}^{EAR}} \times (EvolTaille\ Mén_Prévue - EvolTaille\ Mén_réelle) \right)$$

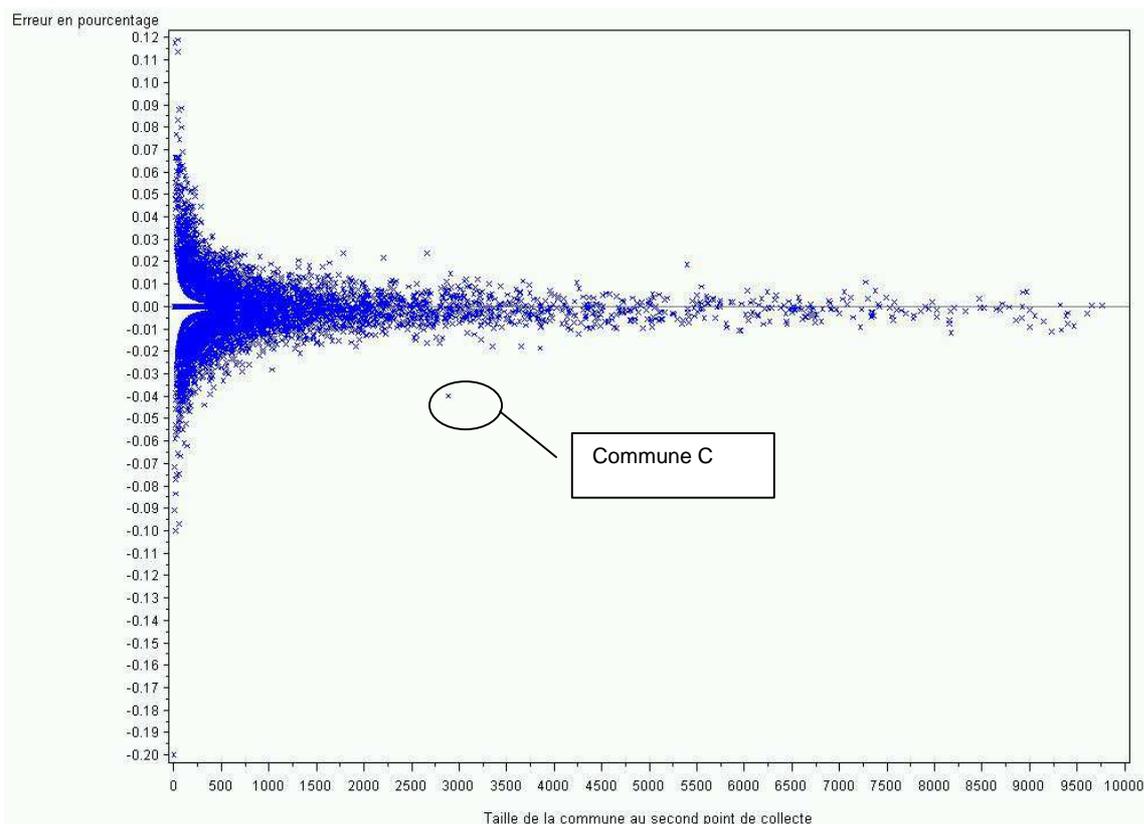
- erreur par rapport à une simple interpolation : l'erreur d'estimation est due également à l'erreur que l'on fait en utilisant la TH pour faire évoluer le nombre de logements. On s'intéressera brièvement à l'erreur « totale » d'estimation effectuée par rapport à une simple interpolation linéaire de la population entre les deux points de collecte, ceci afin d'estimer approximativement la part relative dans l'erreur totale de chacune des estimations (TH et décohabitation). On ne fait donc pas d'hypothèse particulière ici sur l'évolution TH, ou sur la tendance d'évolution de la taille des ménages, on mesure l'écart entre l'estimation réalisée et une interpolation simple de la population.

Par la suite, on s'intéressera principalement au premier type d'erreur, l'erreur à évolution TH inchangée, afin d'essayer d'isoler l'effet de l'erreur de prévision en utilisant une tendance d'évolution de la taille des ménages calculée sur des enquêtes passées.

On peut distinguer les erreurs, selon leur importance relative (à la taille de la commune) ou leur importance en valeur.

Les erreurs relatives (en pourcentage de la population communale) sont liées à la taille de la commune, les erreurs relatives étant, logiquement, plus importantes pour les petites communes (cf. Graphique E) : elles sont le reflet d'une plus forte dispersion du paramètre d'évolution de la taille des ménages et de phénomènes de nature plus conjoncturelle au niveau de ces petites communes.

Graphique E : Erreur de prévision relative selon la taille de la commune



Lecture : Chaque petite commune est représentée par un point. Le nuage de point fournit la relation entre la taille de la commune en abscisse et l'erreur de prévision relative par rapport à la taille des ménages observé entre le prolongement pendant cinq ans de la tendance passée et la valeur observée a posteriori à la collecte du second cycle de recensement. Pour la commune de l'Aube identifiée par le point C, la tendance d'évolution de la taille des ménages mesurée entre 1999 et 2006 correspond à une diminution de la taille des ménages de 4 %. Elle est utilisée comme prévision de l'évolution 2006-2010. L'évolution réelle observée entre 2006 et 2010 est nulle. L'écart entre la prévision et l'évolution réelle implique une sous-estimation de 118 habitants lors de la première extrapolation (2007).

Source : RP 1999, EAR 2004-2013.

Champ : petites communes de France métropolitaine estimées par extrapolation.

En phase d'extrapolation, le calcul de la population municipale d'une commune combine l'effet de l'évolution de la TH et de la tendance d'évolution de la taille des ménages. D'un point de vue global, c'est la TH qui amène les évolutions les plus significatives. En revanche, pour les communes de moins de 100 habitants, le parc de logements est relativement inerte et c'est le prolongement de tendance passé de la taille des ménages qui apporte la plus grande contribution dans le calcul de la population de ces communes (cf. Graphique H partie 4.2). C'est en effet pour ces plus petites communes que sont concentrées les tendances d'évolution de taille des ménages les plus extrêmes. C'est donc naturellement pour ces communes que le risque est le plus important de commettre des erreurs de prévision conséquentes. Ces communes étant également celles pour lesquelles la tendance est la moins pérenne, on y trouvera les erreurs de prévision relatives les plus importantes en fréquence et en niveau.

Pour autant, bien que ces erreurs soient importantes pour les communes concernées, elles ne pèsent que très peu dans les estimations totales de population (cf. Tableau 10). Ainsi, pour les plus petites communes où se concentrent les erreurs les plus significatives relativement à la population de la commune, on ne commet pas d'erreur systématique. A un niveau agrégé, les erreurs commises pour les communes de moins de 100 habitants ont tendance à se compenser : la prévision de la population de l'ensemble de ces communes est tout à fait satisfaisante. On serait donc dans ces communes sur un phénomène d'évolution assez constant dans le temps au niveau global, mais qui est assez aléatoire et n'a pas de raison de se prolonger au niveau communal. Au contraire, dans les communes de taille plus importante, une erreur de prévision de la tendance, même faible relativement à la taille de la commune, conduit à des erreurs en valeur plus importantes pour la population extrapolée. Le prolongement de la tendance passée conduit à sous-estimer la population de ces communes de

manière d'autant plus importante que la taille des communes augmente. Cette sous-estimation traduit un ralentissement du phénomène de diminution de la taille des ménages dans ces communes.

Tableau 10 : Cumul des erreurs commises en fonction de la taille de la commune

Tranche de taille	Nombre de communes	Erreur totale (en nombre d'habitants) à évolution TH inchangée		Erreur moyenne (en nombre d'habitants) à évolution TH inchangée	
		en première extrapolation	en seconde extrapolation	en première extrapolation	en seconde extrapolation
Moins de 100	3 865	474	-463	0,1	-0,1
100 à 249	8 804	-551	-9 008	-0,1	-1,0
250 à 499	7 748	-2 620	-19 471	-0,3	-2,5
500 à 999	6 451	-4 309	-34 007	-0,7	-5,3
Plus de 1000	7 462	-27 254	-183 011	-3,7	-24,5
Total	34 330	-34 260	-245 960	-1,0	-7,2

Lecture : Pour les 3 865 communes de moins de 100 habitants, le cumul des erreurs liées à la différence entre le prolongement de tendance passée de la taille des ménages et la nouvelle tendance calculée entre les deux premiers cycles du nouveau recensement représente 474 personnes, soit 0,1 % de la population de l'ensemble de ces communes.

Source : RP 1999, EAR 2004-2013.

Champ : petites communes de France métropolitaine estimées par extrapolation.

Les erreurs relatives sont plus simples à anticiper, car souvent liées à des évolutions extrêmes. On peut alors essayer de comprendre les phénomènes qui expliquent la tendance d'évolution de la taille des ménages, et éviter de reproduire cette tendance s'ils semblent n'être que ponctuels.

L'analyse de la tendance d'évolution de la taille des ménages après 11 collectes de recensement met en évidence les difficultés liées à la mesure (et la reproduction artificielle) de phénomènes ponctuels, en particulier dans les communes les plus petites. Pour autant, la diminution de la taille des ménages est une tendance générale, que l'on ne peut omettre dans l'extrapolation de la population.

Pour chacun des paramètres d'estimation, il convient donc d'une part d'identifier au mieux les cas particuliers les plus flagrants nécessitant une estimation différente pour éviter de produire des estimations de population qui ne seraient pas crédibles. D'autre part, il est aussi important de voir dans quelle mesure on peut améliorer de façon générale l'extrapolation des petites communes en proposant des scénarios alternatifs jouant conjointement sur les deux paramètres de l'estimation actuelle.

4. Améliorations de la méthode d'extrapolation

Chaque composante utile au calcul des populations des communes dans la phase d'extrapolation a un apport important. Chaque année, elles permettent à ce titre de retranscrire au mieux les évolutions les plus récentes afin de fournir les populations légales de l'ensemble des communes. Néanmoins, l'utilisation des données de taxe d'habitation pour estimer une évolution du nombre de logements comme le prolongement de tendance de la taille des ménages affichent certaines limites justifiant que l'on se pose la question d'une amélioration de la méthode employée actuellement.

4.1. Traitements des cas aberrants

Lors de la phase d'extrapolation, chacune des sources intervenant dans les calculs des populations légales est expertisée afin de déceler d'éventuelles conséquences importantes sur l'évolution de la population communale. Cette expertise permet d'isoler les cas les plus affectés et de proposer si besoin une correction de la source lorsque l'évolution obtenue est incorrecte.

Sont présentés dans l'encadré suivant deux cas présentant des particularités notables de communes de Gironde.

4.1.1. Les séries de logements TH sont corrigées pour 1 % des communes en extrapolation

Les fichiers de la TH fournissent des informations précieuses pour mettre à jour les chiffres de population. Pour autant, la DGFIP qui transmet ces données à l'Insee a un objectif qui ne rejoint pas forcément ceux de l'Insee. Ainsi les services fiscaux peuvent modifier leurs pratiques de gestion ou assurer un niveau de qualité plus faible pour les locaux qui ne sont pas soumis à une taxe d'habitation (certains logements vacants de petites communes par exemple). L'utilisation de cette source à des fins statistiques requiert donc de neutraliser les évolutions provenant de l'activité des services fiscaux plutôt que la réalité du terrain. La détection de cas atypiques nécessite une expertise manuelle. En effet, l'existence de fortes évolutions retracées par la TH ne signifie pas forcément que les données contiennent des anomalies à corriger et ne renseigne pas sur la nature de la correction qu'il convient d'apporter.

Plusieurs contrôles sont ainsi mis en œuvre (forte évolution de la série, différentiel d'évolution détectée entre la série des résidences principales et la série des logements vacants, suspicions de gestion différente des services fiscaux ...) afin de détecter des évolutions atypiques, qui ne reflètent pas l'évolution de la population. Seuls les cas dépassant un seuil sont analysés. Pour assurer un niveau de qualité suffisant au niveau global et au niveau communal, ce seuil est exprimé relativement à la taille de la commune, en élevant les exigences pour les plus grandes communes.

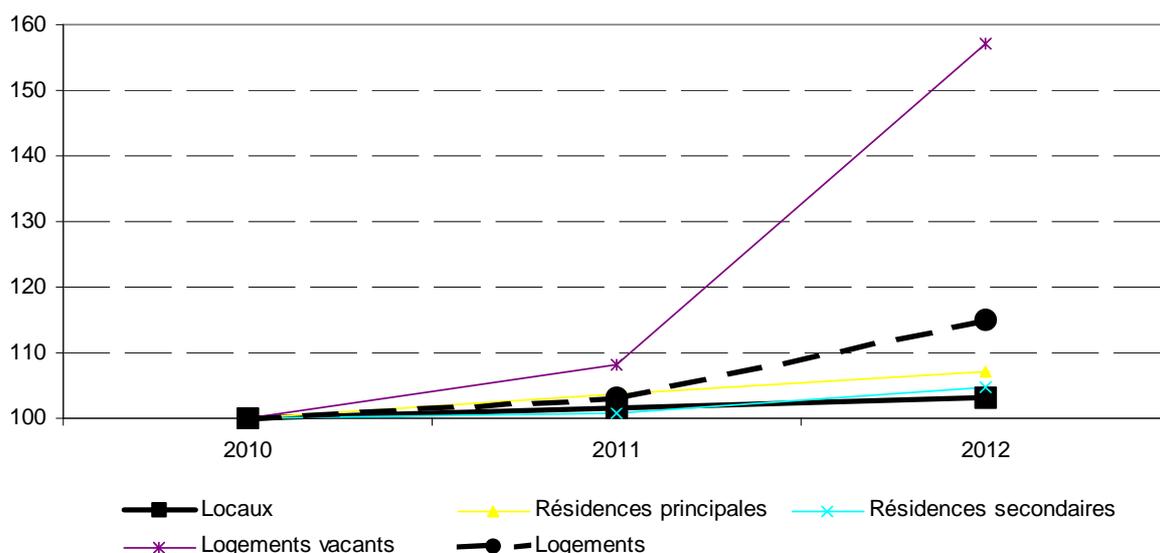
Ces contrôles conduisent à corriger 150 à 200 communes en extrapolation chaque année.

Encadré 2 : Exemple de correction apportée pour les données TH

A titre d'illustration, nous pouvons examiner le cas de la commune A qui a été expertisée dans le cadre de l'extrapolation menant au calcul de population du RP 2012. Cette commune avait été recensée en 2010. Elle était donc estimée par extrapolation, utilisant notamment les informations de la TH pour la deuxième fois au RP 2012.

Cette commune ressort au titre de plusieurs contrôles, et en particulier celui portant sur des évolutions différenciées du nombre de locaux et du nombre de logements (cf. Graphique F).

Graphique F : Évolutions comparées des séries TH de la commune A en base 100 pour l'année de collecte (2010)



Lecture : A la TH, le nombre de logements de cette commune recensée en Gironde a progressé de 15 % alors que le nombre de locaux n'y avait progressé que de 3 %.

L'ensemble des locaux comprend les logements (résidences principales, secondaires et logements vacants), ainsi que certaines dépendances de ces logements (parkings...).

Source : TH 2010-2012.

L'analyse de cette commune a mis en lumière l'apparition de logements de villages vacances introduits dans la base TH entre 2011 et 2012. Ces villages vacances représentent 305 logements qu'il n'y a pas lieu de prendre en compte lors de l'extrapolation puisqu'ils n'apportent pas de population supplémentaire dans la commune. Ils sont donc retirés de la série de logements.

4.1.2. La tendance d'évolution de la taille des ménages est corrigée en cas de choc sur la structure du parc de logements

L'hypothèse sous-jacente à l'utilisation de la « tendance d'évolution de la taille des ménages » est que le nombre moyen de personnes par logement évolue au même rythme sur les 5 années séparant les deux collectes et sur les deux années d'extrapolation.

Cette hypothèse est très forte. En particulier, elle est infirmée pour des petites communes, pour lesquelles les événements démographiques ou économiques sont souvent par nature ponctuels. Par conséquent, il est peu pertinent de rendre ces événements pérennes en utilisant la tendance d'évolution de la taille des ménages sans la contrôler et s'assurer de sa qualité.

Le contrôle sur la tendance d'évolution de la taille des ménages a été mis en place depuis le RP 2011, afin de corriger des situations identifiées a priori comme atypiques, ou reflétant un choc non reproductible.

Le cas de la commune B (cf. Encadré 3) a été détecté au moment de la validation des chiffres de populations légales du RP 2011. Il est apparu opportun de détecter plus systématiquement les communes ayant une situation similaire. Ainsi, a-t-on introduit un contrôle sur la tendance d'évolution de la taille des ménages afin de corriger les évolutions suivant un « choc », en particulier sur le parc de logements de la commune.

Plusieurs types d'événements peuvent engendrer des changements importants, sur la prise en compte des caractéristiques du parc de logement dans les données du RP.

On peut citer en particulier une évolution forte et ponctuelle du parc de logements de la commune, les problèmes de classification des résidences étudiantes, ou des maisons de retraite, qui peuvent être considérées comme des logements ordinaires lors d'une collecte, et comme des communautés lors de la collecte suivante (ou inversement). En raison de leur particularité (ce sont le plus souvent des logements d'une pièce), leur prise en compte dans le parc de logements crée un choc artificiel sur l'évolution de la taille des ménages.

Dans les communes touristiques, la bonne classification des logements (en résidence principale ou secondaire) peut également jouer un rôle important dans l'évolution de la taille des ménages et nécessite une vigilance particulière.

Encadré 3 : Exemple de correction de la tendance d'évolution de la taille des ménages

Le cas de la commune B en Gironde illustre comment une tendance d'évolution de la taille des ménages particulièrement faible conduit à une sous-estimation de la population.

La commune B a été collectée en 2010.

La population de cette commune, tout comme le nombre de logements, a crû fortement entre 2005 et 2010. Pour autant, la tendance d'évolution de la taille des ménages est très faible (-3,2 %) et conduit à estimer à la baisse la population d'une commune qui apparaît par ailleurs comme dynamique.

Tableau 11 : Nombre de logements de la commune B, dans les sources RP et EAR

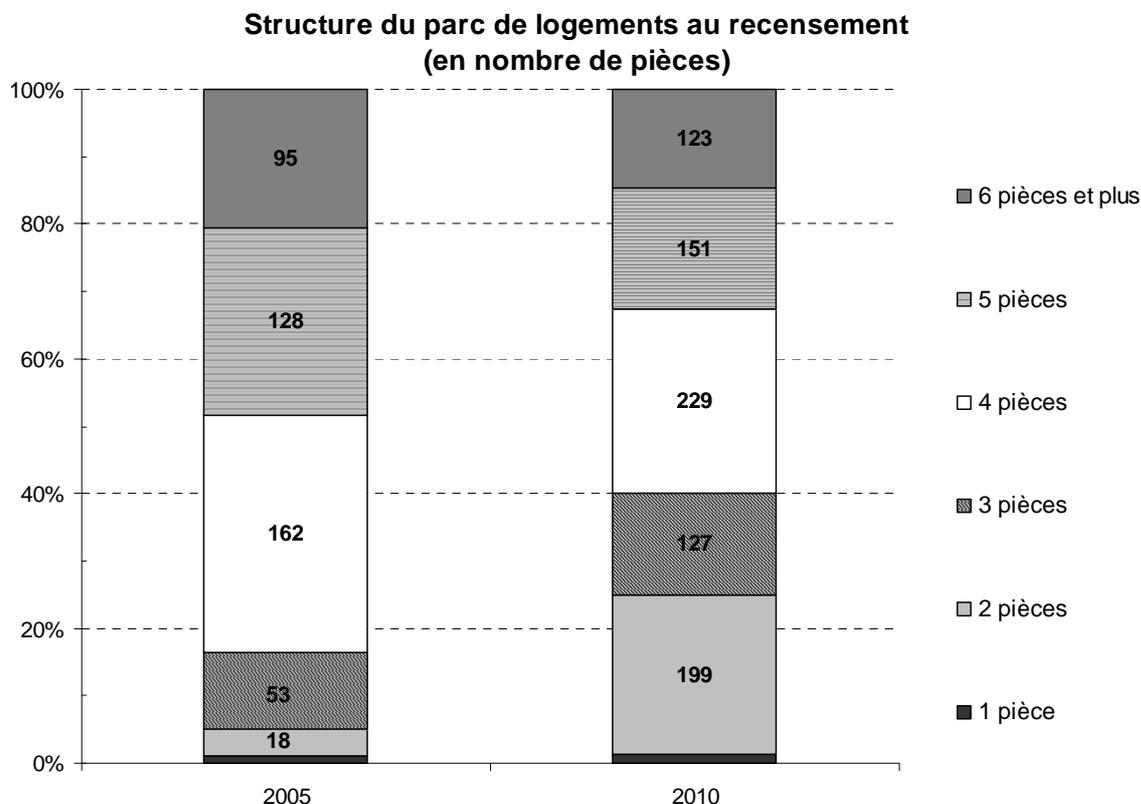
Année	Nombre de logement RP		Nombre de locaux TH	Nombre de logements TH			
	Principaux	Total		Principaux	Secondaires	Vacants	Total
2005 (Collecte)	461	498	530	451	15	36	502
2010 (Collecte)	840	1 021	1 044	766	170	65	1 001
2011 (Extr. 1)	nd	nd	1 071	850	62	117	1 029
2012 (Extr. 2)	nd	nd	1 090	901	21	122	1 044
2013	nd	nd	1 110	899	24	141	1 064

Lecture : Lors de la collecte 2005, la commune B comptait 498 logements. En 2010, elle en comptait 1 021. Les données de la TH indiquent pour la même période une augmentation du nombre de logements (de 502 en 2005 à 1 001 en 2010).

Source : EAR 2005 et 2010, TH 2005, 2010-2013.

Selon la TH (cf. Tableau 11), l'augmentation du nombre de logements après 2010 est nettement plus ralentie que celle de la période 2005-2010. La dynamique urbaine de la commune pour la période 2005-2010 semble donc très différente de la période qui lui a succédé. En analysant plus en détail le parc de logements (cf. Graphique G), on constate une importante déformation du parc de logements entre 2005 et 2010. Les logements de 2 et 3 pièces, très minoritaires en 2005, ont augmenté fortement, en nombre comme en proportion, durant cette période, conduisant à des logements en moyenne plus petits.

Graphique G : Structure du parc de logements de la commune B, lors des enquêtes annuelles de recensement de 2005 et 2010



Lecture : Le nombre de logements de deux pièces est passé de 18 (soit 3,9 %) en 2005 à 199 (soit 23,7 %) en 2010 dans cette commune de Gironde.

Source : EAR 2005, EAR 2010.

L'analyse des données de la TH confirme également la forte croissance du parc de logements de la commune B, ainsi que la déformation de la structure du parc, et le caractère ponctuel de ce phénomène.

Ce profond changement du parc de logements explique la baisse du nombre de personnes par résidence principale sur la période.

Pour le calcul des populations légales de la commune B de 2011 et 2012, si on applique la « tendance d'évolution de la taille des ménages » de la dernière période intercensitaire, alors on suppose que la déformation du parc de logements se poursuit, et donc que le nombre de personnes par logement continue de diminuer au même rythme.

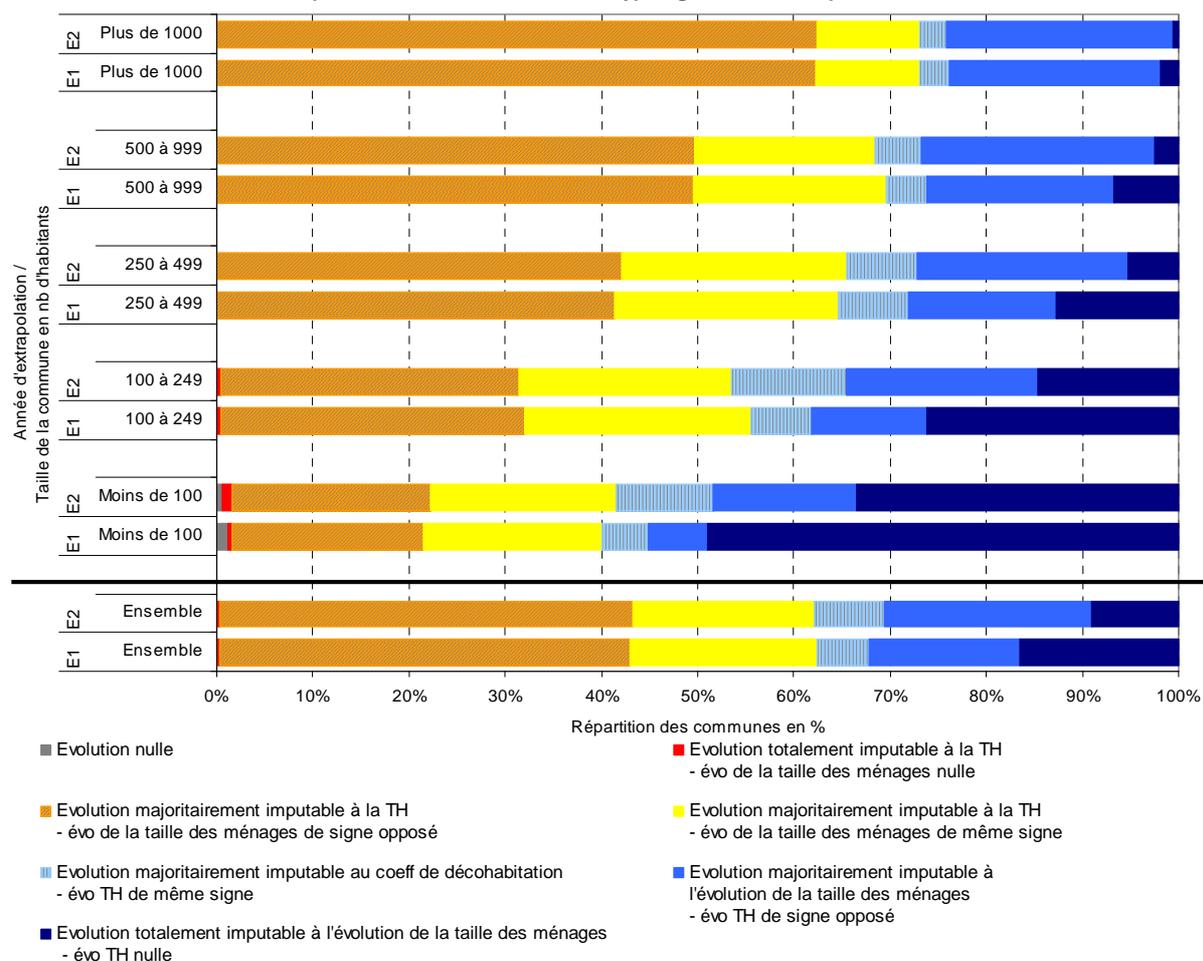
La correction vise donc à prendre en compte le caractère ponctuel de la forte diminution du nombre de personnes par logement, due à la forte modification du parc de logements (ici en faveur des petits logements). On calculera donc une évolution du nombre de personnes par logement, à structure de parc constante, afin de neutraliser l'impact du choc sur le parc de logements.

4.2. Nécessité d'une approche globale pour proposer des améliorations de la méthode d'extrapolation

Les deux ingrédients de l'extrapolation, nombre de logements de la commune et tendance d'évolution de la taille des ménages, ont été analysés séparément. Mais les deux effets peuvent influencer différemment sur les estimations de population, et se compenser mutuellement, tant à un niveau agrégé que pour de nombreuses communes. En effet, l'utilisation actuelle de la source TH conduit plutôt à une surestimation du nombre de logements, et la tendance d'évolution de la taille des ménages aboutit plutôt à une sous-estimation. Pour plus de trois communes sur cinq la surestimation d'un paramètre (souvent la TH) compense la sous-estimation du second (souvent la tendance d'évolution de la taille des ménages). C'est même le cas pour 85 % des communes de 1 000 habitants et plus (cf. Graphique H où ces cas de compensations sont matérialisés par la partie des barres en orange et en bleu foncé). En raison de cette compensation des éventuelles erreurs d'estimation, il est nécessaire de viser à améliorer les deux éléments simultanément, puisque l'amélioration de l'un pourrait conduire à une détérioration de la qualité finale de l'estimation.

L'impact de chacun de ces deux paramètres varie de façon importante, selon la taille de la commune. En effet, dans les communes les plus petites, les constructions ou destructions de logements sont peu fréquentes, et la tendance d'évolution de la taille des ménages sera l'effet prédominant. Pour les communes de moins de 100 habitants, lors de la première extrapolation, l'évolution du nombre de logements TH est nulle pour une commune sur deux.

Graphique H : Répartition des communes en fonction de l'effet dominant pour l'évolution de la population



Lecture : 62 % des communes en première extrapolation de 1 000 habitants et plus (partie orange de la première barre) connaissent une évolution de leur population majoritairement imputable à la TH et dont la tendance d'évolution de la taille des ménages est de signe opposé. Les deux paramètres ne se compensent pas pour seulement 15 % des communes en première extrapolation de 1 000 habitants et plus (parties jaune et bleu claire de la première barre).

En dehors de corrections spécifiques, liées à des phénomènes ponctuels simplement identifiables et corrigibles, les analyses menées jusqu'alors laissent penser que des améliorations sont possibles, de façon plus générale.

Celles-ci doivent bien sûr être construites à partir des éléments dont on dispose. L'analyse a posteriori de la qualité des ingrédients de l'extrapolation nous donne des indications, mais pour la recherche de solutions alternatives, on doit se placer dans les conditions réelles de l'extrapolation où la valeur de la collecte à venir n'est pas encore connue. Ainsi une meilleure utilisation des sources existantes et la plus grande expertise des sources fiscales, utilisées de manière croissante à l'Insee ces dernières années, offrent des perspectives intéressantes à explorer.

Le choix parmi les méthodes alternatives doit donc être validé au travers d'un critère statistique de qualité de l'extrapolation, indépendant de la connaissance de la réalité. Un autre critère d'évaluation de la méthode sera sa simplicité pour en faciliter la compréhension et la communication. En effet, les populations légales des communes et les résultats du recensement de la population présentent un enjeu important pour des acteurs très variés qui ne sont pas forcément rompus à l'utilisation de techniques complexes. Les nombreuses interrogations d'élus quant à l'évolution plus ou moins attendue de la population de leur commune nécessite, par exemple, d'être en mesure d'expliquer simplement la méthode qui est employée par l'Insee.

La comparaison d'une méthode d'extrapolation confrontée au bout de cinq ans à la population qui aura été recensée est une première possibilité. Cette approche déjà retenue pour qualifier la qualité de l'utilisation actuelle des fichiers de la TH permet de fournir une vision globale de la qualité de la méthode mais s'éloigne légèrement de l'emploi de la méthode d'extrapolation qui est en pratique limitée à deux années. Un second critère d'analyse consiste à quantifier le nombre de « retournements de tendance ». Il est complémentaire au premier critère et présente l'avantage de correspondre à un objectif de terrain : une bonne acceptation et appropriation des populations légales par les communes.

On entend par retournement de tendance une rupture dans l'évolution de la population communale au cours du cycle intercensitaire : le plus souvent, l'extrapolation conduit à surestimer la population, et après deux années de croissance, celle-ci décroît lors de la phase d'interpolation. Ce phénomène est mal accepté par les communes lorsqu'il ne reflète pas la réalité perçue.

Par la suite, on présente les différentes pistes envisagées à ce stade.

L'objectif est d'améliorer globalement la qualité des estimations de population communale, lors de l'extrapolation.

4.3. Possibilité d'utilisations alternatives des sources actuelles

Avec la collecte de 2013, on dispose pour la première fois de deux points de collecte pour chaque petite commune. On peut ainsi exploiter les estimations séparant deux points de collecte, pour l'ensemble des quelque 35 700 petites communes métropolitaines. Les trajectoires, en termes de niveau de population, sur les six années incluant deux collectes sont variées. En particulier, on observe un retournement de tendance de la série de population pour près de 53 % des petites communes. Le cas le plus fréquent est une extrapolation trop forte, qui conduit à faire diminuer la population d'une commune lors de la phase d'interpolation. Outre un retournement de tendance, on peut également observer une évolution plus ou moins heurtée, c'est-à-dire s'éloignant d'une évolution linéaire entre les deux collectes.

L'assurance d'une bonne estimation de population dépend d'une part de la qualité de la mise en concordance des données TH avec les besoins du RP, et d'autre part de la pertinence de la méthode d'estimation adoptée. La méthode d'estimation décidée il y a une dizaine d'années explique ainsi en partie l'évolution chahutée de la série de population des ménages pour certaines communes. D'autres estimateurs, dérivés de celui-ci, permettraient notamment de réduire le nombre de communes présentant un retournement de tendance de leur série de population des ménages.

Avec les éléments dont on dispose (TH et recensements précédents), on est en mesure de proposer des estimateurs proches, mais ayant de meilleures propriétés que l'estimateur actuel. Les possibilités d'amélioration du facteur d'estimation du nombre de résidences principales avec la TH d'une part et de la tendance d'évolution de la taille des ménages d'autre part sont nombreuses. Les estimateurs testés combineront des évolutions de chacun des deux facteurs de la méthode l'extrapolation. On peut en outre utiliser la TH de différentes façons, dans l'élaboration du facteur portant sur le nombre de logements comme dans l'élaboration de la tendance d'évolution de la taille des ménages.

Des estimateurs alternatifs peuvent être construits avec des données de TH avancées d'un an, voire lissées sur deux ans. On peut également intervenir sur le champ des logements sur lequel repose le calcul des estimateurs (par exemple, ne pas prendre en compte les logements vacants dont le stock peut être plus ou moins bien mis à jour dans la base TH gérée par la DGFIP).

Enfin, on peut aussi faire des hypothèses sur la tendance d'évolution de la taille des ménages : c'est-à-dire considérer que ces grandeurs proches de 1 sont des constantes, égales à 1 ou à la médiane sur une sous-population des petites communes (stratifiées par taille de commune, par taille moyenne des ménages de chaque commune...). Il serait aussi possible d'allonger la période d'évaluation de la tendance d'évolution de la taille des ménages pour cerner un phénomène plus durable.

Une dernière possibilité d'estimation consisterait à renverser la logique actuelle qui suppose la décomposition de l'évolution en deux grandeurs. La possibilité de proposer une évolution de la population liée à ce que l'on observe au sein des sources fiscales mérite d'être explorée. Elle utiliserait de manière plus complète la source fiscale, ce qui peut représenter un certain risque de dépendance plus grande vis-à-vis d'une source que l'on ne maîtrise pas. En revanche, cette solution aurait l'avantage d'éviter de mesurer une évolution en deux parties dont les concepts et les périodes de calculs sont différents. Elle permettrait également d'éviter de faire l'hypothèse que le taux de résidences principales des communes n'évolue pas pendant les deux années de l'extrapolation.

Chaque proposition d'estimateur alternatif « candidat » pourra être comparée à l'estimateur standard et évaluée aux travers des critères statistiques et de simplicité.

Selon la sensibilité aux différents risques, on pourra retenir l'estimateur qui conduira à la plus grande diminution du nombre de retournements de tendance, celui qui améliore certaines situations, sans en dégrader d'autres, ou faire des distinctions selon que le nouvel estimateur est adapté à toutes les communes, ou plutôt à certaines communes spécifiques (petites ou grandes...).

Une première analyse montre l'intérêt d'adapter la méthode d'estimation appliquée dans le cadre du calcul des populations légales. L'opportunité de changer doit être examinée, à l'aune d'un certain nombre de contraintes pratiques. Les investigations se poursuivent et pourront être prises en compte plus aisément, à un horizon de moyen terme.

5. Conclusion

Les études menées sur l'extrapolation de la population des petites communes ont validé l'innovation du nouveau recensement consistant à recourir à une source administrative, par ailleurs très utilisée par les élus locaux. Elles ont par ailleurs confirmé l'intérêt d'intégrer un paramètre d'évolution de la taille des ménages dans cette phase. La communication entre l'Insee et les communes justifie par ailleurs l'utilisation des deux paramètres estimés au niveau de la commune, ce qui est plus simple pour établir un constat partagé.

Néanmoins, ces analyses ont également mis en lumière les limites et faiblesses de la méthode actuelle. Elles existent pour certains cas particuliers où les hypothèses sur lesquelles reposent l'extrapolation (cohérence entre concept et évolution de la source RP et de la source TH d'une part, pérennité de l'évolution de la taille des ménages d'autre part) ne sont pas vérifiées. Des ajustements ont déjà été mis en place à ce propos. Mais elles sont aussi présentes plus structurellement pour un ensemble de communes partageant le même profil (très petites communes par exemple) dont une correction au cas par cas n'est pas envisageable.

Ainsi en gardant les fondements de la méthode actuelle, investir sur une meilleure utilisation des sources utilisées ouvre des perspectives intéressantes. Elles sont d'ailleurs en cours d'exploration au travers d'autres travaux sur les estimateurs alternatifs permettant la mise en œuvre à terme des améliorations nécessaires pour la production de chiffres de population de meilleure qualité.

Bibliographie

- [1] Desrosières A., « Enquêtes *versus* registres administratifs : réflexions sur la dualité des ressources statistiques », Courrier des statistiques n°111, Insee, 2004.
- [2] Descours L., « Estimation de populations locales par la méthode de la taxe d'habitation », Insee Méthodes n°29-30-31, 1991.
- [3] Direction Générale des Finances Publiques, « Le rapport d'activité 2012 de la direction générale des finances publiques », Tableaux Statistiques pp. 35-54, 2013.
- [4] Godinot A., « La rénovation du recensement de la population », Courrier des statistiques n°105-106, Insee, 2003.
- [5] Godinot A., « Pour comprendre le recensement de la population », Insee Méthodes mai 2005 - N° Hors Série, 2005.
- [6] Héran F., « L'Allemagne et ses minorités ou les surprises du recensement de 2011 », Note d'actualité 5 juin 2013, Ined, 2013.