

**UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE  
CONFERENCE OF EUROPEAN STATISTICIANS  
STATISTICAL OFFICE OF THE EUROPEAN UNION**

**Joint UNECE/Eurostat Expert Group Meeting on Register-Based Censuses  
(The Hague, The Netherlands, 10-11 May 2010)**

Working paper 8  
6 May 2010

**L'utilisation de la taxe d'habitation en France dans la nouvelle  
méthode de recensement.**

**François Clanché**

**Chef du département de la démographie**

Institut national de la statistique et des études économiques (France)

**+33 1 41 17 52 64**

**francois.clanche@insee.fr**

**Résumé**

La France a mis en place, depuis 2004, un recensement tournant. Des enquêtes ont lieu tous les ans sur une partie du territoire. Au début de chaque année, environ 14% de la population remplit un questionnaire de recensement. A la fin de chaque année, un chiffre officiel de population (utilisé dans de nombreux domaines financiers et réglementaires) est calculé pour chaque commune, sur la base des 5 dernières enquêtes. Cette « population légale » est calculée pour chaque commune (il y a 36 882 communes en France), quelle que soit la date de l'enquête sur son territoire.

La méthode est essentiellement basée sur des enquêtes auprès de la population. Mais on utilise des données administratives, celles de la taxe d'habitation, pour ramener à une même date de référence les populations de communes collectées des années différentes. L'objet de cette communication est de présenter les modalités de cette utilisation « statistique » de données administratives.

Quand une commune de moins de 10 000 habitants est recensée un ou deux ans avant la date de référence du recensement, le nombre de logements soumis à la taxe permet une extrapolation de la population. La donnée administrative est donc utilisée en évolution, pour rafraîchir les données de la dernière enquête (partie 1).

La première étape est la collecte, auprès de l'administration fiscale, de données. Elles sont élaborées en fonction de règles et de pratiques qui correspondent à des objectifs fiscaux. Il a fallu comprendre leur mode de construction et leur calendrier de mise à jour.

Une expertise des résultats a ensuite conduit à considérer que l'information qui distingue les résidences principales des autres logements (résidences secondaires, logements vacants), n'était pas suffisamment fiable dans ces bases. La méthode finalement retenue l'évolution du nombre total de logements d'une année sur l'autre. Un coefficient correctif, issu des collectes précédentes du recensement, permet de tenir compte de l'évolution tendancielle du nombre de personnes par logements dans la commune.

Enfin il a fallu établir des règles de « correction » des données quand les évolutions d'année en année sont manifestement erronées, comme des baisses significatives une année suivies de hausses de même ampleur l'année suivant. (partie 2).

Au final, le dispositif fonctionne depuis 2 ans (les populations ont été calculées et publiées en décembre 2008 et décembre 2009) et n'a pas suscité pas de contestations, notamment de la part des 21 000 communes dont la population a été calculée selon cette méthode. Il nécessite cependant des moyens humains permanents non négligeables.

(partie 3).

# 1. Pourquoi utiliser la Taxe d'habitation ?

- La nécessité d'une source annexe aux enquêtes de recensement

Dans la méthode française de recensement, les petites communes (36 000 communes de moins de 10 000 habitants) sont enquêtées de façon exhaustive, par roulement, une fois tous les 5 ans. Leur population est cependant calculée tous les ans, comme celle de l'ensemble des communes. Pour des raisons de cohérence statistique, mais aussi d'équité réglementaire, il est nécessaire que le calcul de la population de chaque commune corresponde à la même année, dite « année de référence ».

Pour des raisons de robustesse statistique, il a été décidé que l'année de référence serait l'année médiane du cycle quinquennal de collecte. Par exemple, les populations publiées à la fin de l'année 2009, sur la base des enquêtes réalisées entre 2005 et 2009, ont pour année de référence 2007.

Pour le cinquième des petites communes enquêtées en 2007, c'est exactement le résultat de l'enquête qui est pris en compte. Pour les communes enquêtées en 2008 ou 2009, on fait l'hypothèse d'une croissance régulière entre le dernier chiffre publié (qui avait comme référence 2006) et l'enquête la plus récente (2008 ou 2009) pour se ramener à l'année 2007 par interpolation linéaire.

En revanche, pour les communes enquêtées en 2005 ou 2006, il faut « rafraîchir » les données pour aboutir à un chiffre 2007. Une simple extrapolation des données sur un ou deux ans, en partant des tendances observées entre les 2 dernières enquêtes, était envisageable. Mais elle comportait des risques importants. En effet ces communes sont petites, voire très petites (la moitié d'entre elles ont moins de 400 habitants). Il n'est donc pas certain que, d'un point de vue démographique, chaque année s'inscrive dans la « tendance » des années précédentes : la « loi des grands nombres » ne s'applique pas sur de si petites valeurs. Dans une commune de 1000 habitants, donc environ 500 logements, on peut voir se succéder deux années avec 5 nouveaux logements (lotissements, petits immeuble) chaque année, et d'autres sans aucune construction. Une simple extrapolation peut donc conduire à des diagnostics erronés, avec sur deux ans, une sur ou sous estimation de la population de plusieurs points.

Avant la mise en place de cette méthode, le dispositif des « recensements complémentaires » permettait de tenir compte des fortes hausses, dans les petites communes, entre deux recensements généraux. Comme les communes en hausse sont plus nombreuses que les communes en baisse, il était essentiel, dans le nouveau dispositif, de maintenir un mécanisme de prise en compte des évolutions entre enquêtes de recensement dans ces communes.

- Les avantages du fichier de la taxe d'habitation

La loi de 2002 qui instaure les nouvelles modalités de recensement prévoit explicitement la possibilité de compléter les données issues des enquêtes d'informations issues de sources administratives.

Nous avons donc cherché une source capable de donner des indications annuelles sur la croissance démographique, fiables au niveau de la commune.

Il n'existe par aujourd'hui de France de base de donnée administrative comportant l'ensemble des habitants avec leurs adresses. En revanche, il existe une source de bonne qualité sur le nombre de logements : le fichier de la « Taxe d'habitation ».

Cette taxe est payée annuellement par tous les habitants des logements. Même si une personne occupe dans l'année plusieurs logements, elle ne la paie qu'une fois. Bien que collectée par l'état, elle est versée uniquement à des collectivités locales, et majoritairement aux communes, pour lesquelles elle représente une source de financement majeure. Les communes sont donc financièrement intéressées à l'exhaustivité et la qualité de cette base. De plus elles sont consultées régulièrement par l'administration fiscale pour vérifier si tous les locaux y figurent. Ceci est un avantage pour le statisticien, car c'est une assez bonne garantie d'exhaustivité, et cela permet de mettre les communes devant leur responsabilité en cas de contestation. La source permet également de distinguer -du moins en principe- les logements occupés en permanence (résidences principales) des autres (résidences secondaires ou occasionnelles, logements vacants).

Par ailleurs, les citoyens souhaitant très rarement être indument imposés deux fois, il y a peu de chances que figurent des doubles comptes.

- Le principe initialement envisagé de son utilisation

Le principe initialement retenu pour l'utilisation de cette source était donc le suivant :

Grâce à l'observation de l'évolution du fichier des logements assujétis à la taxe, on calcul le taux annuel de croissance du nombre de résidences principales entre l'année de la dernière enquête de recensement et l'année de référence. Par ex. 1000 logements dans la TH en 2005, 1020 en 2006, donc  $\frac{20}{1000} = +1\%$  par an. ?

Comme le nombre moyen de personnes par logement n'est pas stable dans le temps, on calcule l'évolution tendancielle du nombre de personnes par logement occupé entre les deux dernières enquêtes de recensement effectuées dans la commune. Par ex -0,5% par an.

On en conclue qu'au total, entre 2004 et 2006, la population a vraisemblablement augmenté de  $(1 - 0,5 = 0,5\%)$  par an, coefficient que l'on applique au nombre d'habitant recensé en 2005 pour aboutir au chiffre de 2007.

La taxe d'habitation est donc dans ce schéma utilisé en évolution, en complément du recensement, mais pas en niveau : le nombre de logement de la TH ne remplace pas celui du RP. Ces données sont ensuite « confirmées » (ou partiellement remises en causes) par l'enquête de recensement intervenant 1 ou 2 ans après le calcul : elles n'ont qu'un impact limité dans le temps sur la population officielle de la commune.

## 2. La mise en place effective du dispositif

Une fois définis ces principes, il faut les mettre en œuvre sur les « vraies » données de la taxe d'habitation. Une convention a été passée entre l'Insee et l'administration fiscale, rendue possible par un arrêté pris après avis de la CNIL (Commission nationale de l'informatique et des libertés). Chaque année, l'administration fiscale fournit à l'Insee un fichier des locaux soumis à la taxe d'habitation.

- Quels locaux prendre en compte parmi ceux figurant dans le fichier fiscal ?

Le fichier que reçoit l'Insee est construit par l'administration fiscale. Il comporte 42 millions d'enregistrements. Il résulte lui-même de la fusion de deux bases :

- une base « foncière », construite essentiellement grâce au cadastre, et qui donne essentiellement une description « physique » des locaux (adressage fin, type de local, taille).
- Une base « contribuables », issue notamment des données de l'impôt sur le revenu, qui donne des informations sur les personnes taxées (nom, nombre de personnes, état civil, adresse du déclarant).

Ces deux bases sont fusionnées par les services fiscaux. Certains locaux ne figurent que dans l'une des deux bases, ce qui constitue en soi parfois un indicateur sur leur nature ou leur occupation réelle.

Dans la base fournie figurent des locaux qui ne sont pas, au sens du recensement, des logements :

- des dépendances (garages, chambres de services, piscines, commerces...) situées à proximité immédiate du logement. Le plus souvent, on peut repérer ces locaux par la combinaison de leurs descriptions « physiques » (issue de la base foncière) et de leur information « contribuable » (ils ont le même déclarant).
- Des locaux meublés à usage d'entreprises ou d'associations. Le plus souvent, l'information sur la nature du déclarant permet de les isoler, car ils ne sont pas assujétis au final à la taxe.

Les bâtiments servant aux exploitations rurales (granges, écuries...). Ici, l'information sur la nature du local permet de les isoler.

Il n'y a pas dans le fichier de variable correspondant exactement au concept statistique de logement. L'Insee construit un « filtre », utilisant plusieurs variables, tantôt d'origine « foncière » tantôt d'origine « contribuable » pour réduire le fichier aux seuls locaux vraisemblablement d'habitation. Ce filtre conduit à supprimer 25% des locaux initialement présents dans le fichier fiscal.

L'Insee construit donc, pour son usage propre, un fichier appelé de la « TH filtré » qui sera utilisé dans plusieurs opérations liées au recensement : mise à jour du répertoire des immeubles localisés (RIL) dans les grandes communes, contrôle de la qualité de la collecte dans les petites communes, calcul de la population par extrapolation dans certaines petites communes. La suite du texte ne portera que sur ce troisième usage.

- La notion de « résidence principale » et sa fragilité.

Parmi les logements ainsi identifiés, seules les « résidences principales », celles qui abritent des personnes une majeure partie de l'année, doivent en principe être prises en compte pour participer à la « mise à jour » des données du recensement. Malgré l'existence d'une telle catégorie dans la source fiscale, il n'est pas simple de les repérer dans ce fichier, et ceci pour plusieurs raisons.

Est considérée comme « principale » pour la taxe d'habitation une résidence où une personne a déclaré ses revenus (au titre de l'impôt sur le revenu) au début de l'année précédente. Or une personne ne déclare pas forcément ses revenus là où elle vit la majeure partie de l'année : elle n'y est pas obligée. Les taux d'impositions à la taxe d'habitation sont variables selon les communes. Ils sont également différents selon qu'un logement est déclaré ou pas « résidence principale ». Ainsi une personne peut avoir intérêt à déclarer ses revenus dans un logement où elle réside seulement une petite partie de l'année. On peut donc avoir, en niveau, des différences notables entre le nombre de résidences principales au sens du recensement et ce même nombre au sens de la TH. En général, les résidences secondaires paient plus d'impôt que les résidences principales : ni les mairies, ni l'administration fiscale n'ont donc intérêt à faire diminuer leur nombre

De plus, le décalage d'un an entre la déclaration de l'impôt sur le revenu et sa prise en compte dans la taxe d'habitation rend délicates les analyses en évolution d'une année sur l'autre, surtout sur de petits effectifs. Une évolution de statut peut aussi provenir d'une « stratégie fiscale » d'un contribuable qui a intérêt à changer de résidence au sens de l'impôt sans pour autant déménager réellement.

Empiriquement, on constate une plus grande volatilité des séries de résidences principales que des séries de logements en général : les évolutions fortes d'une année sur l'autre, les retournements de tendances significatifs sont nettement plus nombreux dans les séries de résidences principales que dans les séries générales de logements.

Il est vite apparu, lors des premiers tests d'applications de la méthode initialement prévue, que ces irrégularités entraînaient un nombre non négligeable d'évolutions très improbables des populations concernées. Pour que l'ensemble des données soient crédibles, il aurait fallu gérer de façon manuelle un grand nombre de cas particuliers, ce qui aurait été coûteux, aurait certainement créé des distortions et des inéquités entre communes et aurait entraîné des contestations, voire des contentieux. Si les communes, dans les années qui viennent, « surveillent » mieux leurs fichiers de TH, la qualité des séries « résidences principales » pourra s'améliorer et on pourra reconsidérer notre position.

- La solution finalement choisie

Il a donc été décidé de mettre en place un dispositif plus robuste : c'est la série de l'ensemble des logements qui est prise en compte pour définir la tendance d'extrapolation sur un ou deux ans, et elle est corrigée par l'évolution tendancielle du nombre moyen de personne par logements (et non pas par résidences principales).

Cette solution a aussi l'avantage que le chiffre du total des logements est mieux connus des communes que celui des résidences principales, et que sa signification est moins contestable par les communes du fait des « stratégies fiscales » des habitants.

Exemple : Monchiffray, commune fictive de moins de 10 000 habitants recensée en 2005

	1999	...	2005	2006	2007
<b>Données de recensement :</b>					
Logements	858		924		
Population	1859		1950		
Nombre moyen de personnes par logement	2,17		2,11		
<b>Données de la taxe d'habitation</b>					
Nombre de logements			910	923	951

Évolutions annuelles moyennes issues des recensements de 1999 et 2005 :

du parc de logements :	1,24%	(1)
de la population :	0,80%	(2)

Différentiel annuel d'évolution ( (1) - (2) )	0,44%	(3)
Évolution annuelle moyenne du parc de logements TH entre 2005 et 2007 :	2,23%	(4)
Progression annuelle moyenne de la population estimée (4 - 3) :		1,78%
Population estimée pour 2007 :	2 020	habitants

- La gestion des évolutions « aberrantes »

Au niveau d'une commune, la série des nombres de logements offre le plus souvent un profil régulier : sur une période de 3 ans, 83% des communes sont en hausse 3 années de suite, 9% sont en baisse continue. Pour environ 7% le profil est irrégulier mais sans renversements de tendance significatifs. De ce fait, l'utilisation des évolutions sur un ou deux ans ne pose pas de problème.

En revanche, pour 0,5% des communes, les profils sont surprenants : une année de « pic » suivi d'un retour sur une tendance régulière, ou une année de « creux » suivi d'un retour à la tendance, ou encore un double retournement.

On a du définir des seuils, forcément différents selon la taille de la commune, à partir desquels on a décidé de « lisser » artificiellement les séries, en faisant l'hypothèse que l'année de pic (ou de creux) étaient dues à des erreurs. Ce seuil a été fixé à 6% des logements pour une commune de 100 habitants, 4% pour 500 habitants, 0,6% pour 5 000, etc...

Chaque année, 300 communes sur 14 000 qui font ainsi l'objet d'un traitement particulier. Il s'agit des corrections systématiques de la source : le même traitement est appliqué à toutes les communes présentant le même « profil ». Ces anomalies apparentes dans les fichiers font l'objet d'un signalement aux services fiscaux, mais nous ne connaissons malheureusement pas toujours les corrections ont pu être apportées par la suite. Chaque année, après la date de l'envoi des fichiers à l'Insee, les communes peuvent signaler des erreurs à l'administration fiscale. Mais les corrections apportées ne sont pas prises en compte dans un fichier centralisé. De ce fait, nous n'avons pas connaissance de ces corrections. C'est regrettable, car la correction de l'année N-1 améliore l'analyse des évolutions.

### 3. Une opération annuelle efficace mais coûteuse

L'ensemble de ces règles de gestion et de traitements ont été contruites et mises au point au cours des années 2005 à 2007, pour une première mise en œuvre en grandeur réelle en 2008, lors du premier calcul des populations de toutes les communes basées sur les enquêtes 2004 à 2008. Le processus est désormais reproduit tous les ans.

- Des opérations de vérification et de correction « centralisées »

Chaque année, les fichiers relatifs à la taxe d'habitation de l'année, correspondant approximativement à la situation du début de l'année, sont mises à la disposition de l'Insee vers le mois de décembre. Plus précisément, ce sont 107 fichiers correspondant aux 107 centres des impôts que compte la France.

L'Insee réceptionne les fichiers, en vérifie l'intégrité formelle et la « vraisemblance » en terme de volume avant de faire passer tous les traitements automatisés décrits précédemment : filtrage, corrections systématiques des cas aberrants. Se rajoutent à ces opérations d'autres traitements liés aux autres usages de la TH, notamment sur les adresses. Leur durée totale est d'environ 2 mois, pour aboutir à des calculs provisoires.

- Des expertises locales sur certains points particuliers

Les calculs, effectués centralement, sont envoyés pour vérifications, aux directions régionales de l'Insee.

L'examen, lors des premières campagnes de calcul, des résultats commune par commune a fait apparaître, malgré les contrôles déjà mentionnés, certaines évolutions fortes d'une année sur l'autre dans des communes pourtant apparemment « stables ». L'analyse de ces situations a en général fait apparaître des phénomènes suivants : l'apparition -ou la disparition- d'une adresse importante dans la

commune est liée au calendrier des procédures administratives mais n'est pas cohérent avec les définitions ou les protocoles du recensements.

Deux exemples :

un immeuble de 20 logements, vétuste, appelé à être détruit, inhabitable car muré, disparaît du fichier de la taxe d'habitation entre l'année 2005 et l'année 2006, car il a été effectivement détruit. Or en janvier 2005, l'agent recenseur n'a pas pris en compte cet immeuble, conformément aux consignes, puisqu'il n'était absolument pas habitable. Le risque est donc diminuer la population de la commune d'environ 40 habitants en 2006 en les soustrayant indument à la population recensée en 2005. Dans ce cas, la solution technique est d'enlever l'immeuble du fichier TH de 2005.

En sens inverse, 25 logements apparaissent dans le fichier de la TH d'une petite commune entre 2005 et 2006. Or à la même adresse, une maison de retraite a été enquêtée comme communauté en 2005, avec 29 personnes. Dans l'intervalle, le statut de la communauté a évolué et ses habitants sont désormais éligibles à la taxe, ce qui n'était pas le cas en 2005. Il serait donc erroné d'ajouter ces 25 logements dans le calcul de l'évolution 2005-2006.

Ces situations spécifiques, rares, ne peuvent être détectées et analysées qu'un cas par cas, dans les équipes régionales. Les équipes centrales « préparent » cependant le travail en reprérant les communes dont les évolutions à la TH sont quantitativement significatives et pour lesquelles une seule adresse a un impact important. Ceci permet de cibler les expertises.

- Une réussite technique et politique

En 2008, puis en 2009, cette méthode de calcul a été appliquée à deux cinquièmes des « petites communes », soient environ 21 000 communes au total, qui ne sont pas les mêmes chaque année. La moitié d'entre elles sont extrapolées sur un an, l'autre sur deux ans.

L'année 2008, première année de publication, lorsque les communes ont reçu le courrier les informant de leur population et de leur mode de calcul, plusieurs centaines d'entre eux ont demandé des explications complémentaires. Mais aucune n'a contesté le résultat, et il n'y a pas de contentieux dans ce domaine. Cet accueil favorable a été rendu possible par une importante communication préalable auprès des communes : les responsables régionaux de l'Insee, ainsi que les équipes en charge du recensement, ont rentré les maires au cours des années 2007 et 2008 et leur ont expliqué le principe de ce calcul et sa justification.

En 2009, une seconde série de calculs a été réalisée, portant sur l'année de référence 2007. On a pu mesurer des évolutions par rapport au point précédent. On a pu également constater, à posteriori, la qualité de la « prévision » par la taxe d'habitation des données de l'année 2006 pour les communes qui ont été enquêtées en 2009 (après l'avoir été en 2004). Globalement, cette seconde année n'a pas fait apparaître de problèmes non prévus, et n'a pas conduit à remettre en cause la méthode.

Alors même que, pour la première fois, elles recevaient des populations « officielles » deux années de suite sans avoir réalisé deux enquêtes de recensement, les communes ont été peu nombreuses à réagir lors de la communication de ces chiffres.

- Le coût du dispositif

La mise en place de ce dispositif a nécessité un long travail de discussion avec l'administration fiscale, d'expertise de la source, d'analyse sur des cas particuliers, de tests sur des séries de données, qui a impliqué des cadres supérieurs méthodologues à la direction générale de l'Insee ainsi que des agents en charge de la réalisation du recensement et du dialogue avec les communes dans les établissements régionaux. On peut estimer le coût « statistique » du projet à l'équivalent de 4 années/personne de méthodologues de bon niveau. Il faut y ajouter le développement des outils informatiques capables d'accueillir les sources et de les traiter, soit environ 1 an d'informaticien.

Bien que l'opération n'en soit qu'à sa troisième réalisation en 2010, on peut considérer qu'il faut un cadre statisticien à temps complet sur l'année pour gérer ce dispositif en régime de croisière, auquel s'ajoute l'équivalent d'une autre personne réparti sur des équipes régionales, informatiques ou d'appui statistique central. A l'avenir, une meilleure mutualisation des usages de ce fichier au sein de l'Insee (il y en a plusieurs, dans des domaines différents), pourrait permettre de réduire ce coût.