

B. RIANDEY

M. H. GENSBITTEL

**Séminaire méthodologique SFdS-INSEE sur la
rénovation du recensement de la population RRP.
Actes de la séance du 24 octobre 2001**

Journal de la société française de statistique, tome 142, n° 3 (2001),
p. 3-52

http://www.numdam.org/item?id=JSFS_2001__142_3_3_0

© Société française de statistique, 2001, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

Séminaire méthodologique SFdS-INSEE sur la rénovation du recensement de la population (RRP)

Actes de la séance du 24 octobre 2001

Introduction

Le recensement général de la population est un élément majeur de la production du système statistique public, par sa portée symbolique et par ses considérables apports dans les domaines de la connaissance et des travaux statistiques : le recensement est à la source de la définition de la population légale, fournit des données irremplaçables pour de nombreux travaux de recherche sur la population française et ses caractéristiques, et apporte données de cadrage et bases de sondage indispensables aux statisticiens.

Il est donc normal que la SFdS, représentant aussi bien la communauté scientifique que les statisticiens professionnels, manifeste un fort intérêt pour la rénovation du recensement, et contribue au nécessaire dialogue entre l'INSEE et les utilisateurs du recensement.

Les textes qui sont ici présentés retracent les exposés et débats qui ont eu lieu lors d'un séminaire de réflexion organisé conjointement par l'INSEE et la SFdS à l'automne 2001, un an après un premier séminaire dont les actes ont été publiés dans le tome 140 n° 4 du journal de la SFdS.

Entre ces deux dates, l'INSEE a pu procéder à des expérimentations qui l'ont conduit, on le verra, à modifier sensiblement la méthode de collecte dans les zones urbaines.

Enfin, les nombreux utilisateurs du recensement qui trouvent en lui une source unique de données localisées ont pu prendre connaissance des méthodes de construction des estimations détaillées mises à l'étude par l'INSEE, de la nature et de la qualité des données recueillies, et du calendrier de leur mise à disposition. Soulignons qu'une avancée importante a été observée quant au traitement de la mobilité géographique et des données bilocalisées.

Compte tenu des enjeux, les débats ont été animés et constructifs, mais ne sont manifestement pas clos : un nouveau séminaire permettra sans doute de les poursuivre prochainement.

B. Riandey
M.H. Gensbittel

LA RÉNOVATION DU RECENSEMENT DE LA POPULATION

Alain GODINOT *

En préambule de cette présentation des actualités de la rénovation du recensement de la population, je voudrais remercier la Société française de statistique et ceux de ses membres qui participent activement au bon déroulement du séminaire SFdS-INSEE, et notamment Messieurs Gensbittel et Leridon

1. Les actualités d'ordre juridique

Ces actualités sont, en premier lieu, de nature juridique. Les articles de la loi concernant la rénovation ont été insérés dans le projet de loi sur la démocratie de proximité, en cours de discussion au Parlement. Ils ont été votés en première lecture à l'Assemblée nationale le 25 juin 2001 et enrichis d'un amendement visant à créer une commission rassemblant des représentants de collectivités locales, des représentants de l'État et des statisticiens pour déterminer les modalités des enquêtes par sondage.

Cet amendement offre une bonne opportunité d'approfondir la concertation en cours. Il serait souhaitable, du point de vue de l'Insee, que la commission soit constituée au sein du Conseil national de l'information statistique (CNIS), dont c'est la mission d'examiner l'opportunité et les caractéristiques techniques des opérations de la statistique publique, et qui a organisé et mené depuis 1999 une concertation statistique large sur le programme de rénovation. La première lecture au Sénat est annoncée pour janvier 2002. Le gouvernement ayant décrété l'urgence sur le projet de loi « Démocratie de proximité », ce texte devrait être adopté avant la fin de la présente législature, au printemps 2002¹. Ce calendrier permettrait aux premières enquêtes de recensement de démarrer début 2004, les premières populations légales issues du nouveau système étant alors publiées fin 2008.

* INSEE, programme de rénovation du recensement de la population, 18, boulevard A. Pinard, 75675 Paris Cedex 14.
E-mail : alain.godinot@insee.fr

1. Après l'examen en première lecture au Sénat le 24 janvier 2002, l'adoption du projet en Commission mixte paritaire et les secondes lectures à l'Assemblée nationale et au Sénat, la loi n° 2002-276 a été promulguée le 27 février 2002.

2. Les inflexions liées aux tests et aux expérimentations

L'actualité de la rénovation du recensement de la population, ce sont aussi les inflexions que nous sommes conduits à réaliser à l'issue des réflexions et des tests. Il en sera question dans la suite du séminaire. Je voulais rappeler en introduction la décision qui a été prise concernant le rythme de la collecte. Celle-ci sera concentrée sur une seule vague annuelle d'enquêtes de recensement, en janvier et février. L'étalement des enquêtes de recensement, initialement envisagé de septembre à juin, aurait posé de difficiles problèmes techniques (comment ramener à une même date des informations collectées tout au long de l'année?) et politiques (comment répartir équitablement les communes pour la collecte dans un calendrier annuel?).

Toujours dans le fil des expérimentations, depuis le séminaire d'octobre 2000, l'Insee a mis en place, à la direction régionale de Rhône-Alpes, un site expérimental dont la mission sera de tester, avec des communes volontaires, les procédures du recensement rénové. Le site commence à travailler avec les communes de Lyon (pour le 5^e arrondissement), de Villefranche-sur-Saône et de Anse. Ses premières tâches consisteront à définir avec ces communes les procédures de mise à jour du répertoire d'immeubles localisés (à Lyon et Villefranche-sur-Saône) et les procédures de réalisation des enquêtes de recensement (dans les trois communes).

3. La concertation

L'actualité comporte aussi de nombreuses actions de concertation. Le dialogue avec les communes se poursuit au niveau local. Six directions régionales de l'Insee examinent avec des élus et des techniciens les modalités concrètes de la préparation et de la réalisation des enquêtes de recensement. Ce dialogue de terrain sera élargi début 2002 à l'ensemble des directions régionales et il nourrira la concertation qui, au niveau national, se tient entre les associations d'élus (au premier rang desquelles l'Association des maires de France), le ministère de l'intérieur (direction générale des collectivités locales) et l'Insee.

Le processus de concertation sur le contenu des principaux questionnaires du recensement (la feuille de logement et le bulletin individuel) arrive à son terme. Le groupe du CNIS créé en 2000 examinera le 5 novembre 2001 les propositions de la maîtrise d'ouvrage et formulera ses ultimes propositions². Après quoi le comité directeur du RRP arrêtera, le 28 novembre³, le contenu de ces deux imprimés, qui devrait rester stable pendant le premier cycle de cinq ans.

2. Le compte rendu de cette réunion est consultable sur le site du CNIS (<http://www.cnis.fr>)

3. Le comité directeur a intégré dans ses arbitrages les priorités données par le groupe du CNIS quant aux questions à poser tout en rappelant l'importance à accorder aux critères de simplification et d'acceptabilité pour tout questionnement et tout particulièrement pour le recensement.

LA RÉNOVATION DU RECENSEMENT DE LA POPULATION

Par la suite, une concertation s'ouvrira sous l'égide du CNIS sur le thème de la diffusion des résultats du recensement rénové. Elle sera éclairée par un bilan de la diffusion des résultats du RP99, qui devrait faire l'objet d'une Rencontre du CNIS à l'automne 2002. Après cette Rencontre, un groupe de travail du Conseil devrait être constitué pour réfléchir aux produits et aux supports de diffusion des résultats du recensement rénové.

La concertation passe aussi par le dialogue avec la communauté scientifique. Le séminaire SFdS-Insee du 5 octobre 2000, dont celui de ce jour est le prolongement, s'inscrit dans cette démarche. De même, l'Insee se fait un devoir de participer aux séminaires auxquels il est convié pour expliquer les tenants et aboutissants de la rénovation du recensement de la population. C'est ainsi par exemple qu'il a participé au séminaire sur l'évolution des recensements organisé en juin 2001 par le LASMAS-Institut du longitudinal.

Enfin, un comité de conseil méthodologique a été mis en place début 2001, auquel participent plusieurs experts statisticiens. M. Jacques Antoine a bien voulu accepter d'assurer la présidence des réunions qui se tiendraient à la demande de la maîtrise d'ouvrage si les procédures d'échanges écrits (qui constituent le principal mode de fonctionnement du comité) appelaient un complément d'échanges oraux.

En conclusion, il reste deux ans à l'Insee pour monter un dispositif de collecte robuste et fiable; beaucoup reste encore à faire et j'attends des échanges au cours de ce séminaire et des réactions à la publication de ces actes des éclairages pour définir au mieux les priorités dans ces travaux préparatoires.

LES TESTS DE COLLECTE MENÉS EN 2001

Résultats et décisions

Jean-Michel DURR *

Le programme de la rénovation du recensement est entré depuis l'an dernier dans une phase de tests plus intense pour préparer la mise en production de l'opération. Au cours du premier semestre 2001, plusieurs opérations ont été menées avec pour objectifs de tester les questionnaires, vérifier la faisabilité technique du protocole de collecte et comparer les taux de réponse selon le mode de sondage.

1. Les tests des questionnaires

Les tests des questionnaires avaient pour but d'évaluer les réactions des habitants face aux différentes formes proposées et d'apprécier la facilité d'auto-administration, et en particulier la compréhension des questions. Des tests qualitatifs ont été conduits selon la méthode des groupes de discussion. Il en ressort que la forme cahier, qui consiste à regrouper les questionnaires d'un même ménage dans un cahier, est moins appréciée, car jugée trop lourde sauf par les ménages nombreux. La forme traditionnelle, constituée de bulletins individuels en feuillets séparés, regroupés dans une feuille de logement servant de chemise, est préférée car plus légère. De plus, ces tests ont permis de recueillir de nombreuses suggestions sur le cheminement dans le questionnaire (place des exemples, formulation des filtres, etc.).

2. La faisabilité technique du protocole de collecte

Les tests quantitatifs ont été réalisés auprès de 4 000 logements dans trois directions régionales : Franche-Comté, Lorraine et Haute-Normandie. Deux modes de sondage ont été expérimentés, le mode « liste », qui consiste à effectuer un pré-recensement afin de dresser la liste des logements, à en tirer 40 %, puis à confier à un agent recenseur la liste et la localisation des logements à enquêter ; et le mode « adresses », qui conduit à enquêter tous les logements d'une adresse donnée. 2 000 logements ont été recensés dans chacun des modes,

* INSEE, programme de rénovation du recensement de la population, 18, boulevard A Pinard, 75675 Paris Cedex 14
E-mail . jean michel.durr@insee fr

LES TESTS DE COLLECTE MENÉS EN 2001

chaque agent recenseur ayant à travailler pour moitié selon chacun des deux modes afin d'éliminer les effets liés à l'enquêteur. Le test avait été déclaré d'intérêt général par le comité du label et non obligatoire. Un dispositif de communication sommaire a été mis en place : affichettes, courriers et avis de passage. Les agents recenseurs ont été recrutés pour l'occasion et n'avaient pas d'expérience particulière des enquêtes.

Le pré-recensement s'est déroulé en mars, la collecte en mai-juin et les contrôles en juin-juillet. Les principales conclusions sont les suivantes :

- la localisation des adresses s'est déroulée convenablement, seulement un peu moins de 1 % d'adresses imprécises ont été trouvées. Ces cas peuvent s'expliquer par le fait qu'on utilisait les informations brutes du recensement de 1999, avant géocodage et affinage ;
- le pré-recensement est une opération complexe et délicate à mettre en œuvre :

1. Il est difficile de pénétrer dans les immeubles au seul motif d'en dresser la liste des logements avec le nom des occupants. De nombreux immeubles munis d'un digicode ou d'un interphone nécessitent un contact avec au moins un occupant pour y pénétrer. Au total, 7 % des adresses n'ont pu être pré-recensées avant intervention d'un agent de l'Insee ;

2. La localisation du logement à l'aide d'informations descriptives (situation sur le palier, etc.) s'avère difficile à utiliser par l'agent recenseur, et les schémas ne l'aident guère. La localisation des logements s'effectue principalement à l'aide du nom, ce qui n'est pas satisfaisant, sachant que le nom n'est présent que dans 60 % ou 70 % des cas, qu'il peut changer entre le moment du pré-recensement et la date de la collecte et qu'il peut y avoir des homonymes dans l'immeuble ;

3. La saisie des informations du pré-recensement, le tirage et la restitution de la liste des logements à enquêter posent de redoutables problèmes d'organisation. La saisie du carnet papier s'avère difficile en raison d'une écriture peu déchiffrable de l'agent pré-recenseur, qui le remplit le plus souvent debout dans l'immeuble ;

- la collecte de type liste pose également des difficultés :

1. L'argumentaire vis-à-vis des personnes enquêtées est plus délicat : il faut répondre à la question « pourquoi moi et pas mon voisin ? » ;

2. L'obtention de renseignements sur le caractère vacant ou non d'un logement à recenser est également plus difficile auprès des voisins non enquêtés. Les agents recenseurs ont d'ailleurs largement préféré le mode « adresses » et l'ont trouvé plus facile à réaliser ;

3. Le taux de réponse est un peu plus faible en mode liste, de l'ordre de 3 points, ce qui est insuffisant pour justifier sur ce seul élément un rejet du mode « liste » mais cela constitue un élément négatif supplémentaire.

La qualité de l'information collectée est plutôt bonne. Les contrôles effectués ont mis en évidence un taux de 1,2 % d'erreur sur la catégorie de logement (principal, vacant ou secondaire), ce qui est comparable aux résultats de

l'enquête post-censitaire de 1990. La répartition des occupants habituels ou non à l'aide des listes A, B et C de l'intérieur de la feuille de logement présente également assez peu d'erreurs.

Concernant les outils de collecte, les tests ont fait remonter des propositions d'amélioration des carnets de tournée, et les listes fournies ont été appréciées. La fiche d'enquête non aboutie (FENA), que l'on souhaite substituer aux « bulletins pour ordre » des recensements précédents, s'avère un outil utile, même si des améliorations sont à prévoir et son utilisation à encadrer. Elle permet un contrôle des « impossibles à joindre » et des « refus » déclarés par les agents recenseurs.

Les tests ont également permis de préciser les procédures de contrôle et de roder notamment le rapprochement avec le fichier de la taxe locale d'habitation.

En conclusion, le mode liste apporte une réduction de la variance mais engendre un biais de couverture non maîtrisable. La précision globale s'en trouve donc affectée, dans une mesure difficile à apprécier, mais qui semble de nature à fragiliser le processus de collecte. La décision a donc été prise de privilégier le tirage d'adresses complètes afin de renforcer la robustesse du dispositif et de faciliter les opérations de collecte. Le plan de sondage sera optimisé dans cette direction. Par ailleurs, ces tests ont confirmé la bonne acceptabilité des questionnaires individuels en feuillets séparés, regroupés à l'intérieur d'une feuille de logement servant de chemise aux bulletins d'un même logement. En conséquence, il a été décidé de conserver cette forme habituelle.

Au cours des mois à venir, d'autres tests seront menés. En novembre 2001¹, 7 000 logements appartenant à 11 communes de plus de 10 000 habitants seront interrogés afin de finaliser le protocole de collecte et notamment la tournée de reconnaissance des adresses à recenser, la pré-numérotation des questionnaires et le suivi de collecte. Au cours de l'année 2002, des tests auront également lieu dans des communes de moins de 10 000 habitants. De plus, comme l'a précisé plus haut Alain Godinot, le site expérimental constitué auprès de la direction régionale de l'Insee de Rhône-Alpes nous permettra d'approfondir avec quelques communes volontaires la répartition des rôles entre l'Insee et les communes et de tester les organisations à mettre en place. Ce site mènera des tests de collecte et d'organisation au cours de l'année 2002.

1. Au total, dans les 11 communes concernées, 4 600 logements ont été recensés dans 2 000 adresses. Les tests ont conforté le mode opératoire prévu pour la tournée de reconnaissance des adresses en début de collecte et la pertinence d'indicateurs hebdomadaires pour le suivi de la collecte. Les contrôles comparant le nombre de logements entre le fichier de la taxe d'habitation et le recensement ont montré que pour 91,5 % des adresses les différences étaient minimes.

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

Exploitation du test RRP mené en mai-juin 2001

Barbara CHRISTIAN *

Nous présentons aujourd'hui des résultats de tests relatifs à l'influence du mode de sondage sur la réponse des ménages. Il s'agissait de comparer les taux de réponse observés dans un sondage avec tirage de logements et dans un plan de sondage avec tirage d'adresses, tous les logements d'une adresse étant enquêtés.

1. Le test

Le test s'est déroulé sur 48 IRIS (quartiers) dans huit communes, réparties dans trois régions : Lorraine, Franche-Comté et Haute-Normandie.

Les IRIS ont été préalablement classés par type de bâti, prenant en compte la variété de l'habitat pour la collecte. Quatre types ont ainsi été définis : IRIS à dominante « immeubles anciens », IRIS à dominante « habitat individuel », IRIS à forte concentration de « gros immeubles » et IRIS « hétérogènes ». Idéalement, chacun des quatre types d'IRIS devait être représenté proportionnellement à son poids dans l'ensemble des 48 IRIS.

Pour la moitié des IRIS (soit 24 IRIS), la collecte a été réalisée selon le mode de sondage suivant : on tire au hasard 1/5 des adresses, que l'on prérecense, c'est-à-dire dont on dresse la liste des logements, puis on sélectionne 40 % des logements ainsi répertoriés dans ces adresses, et on fait passer, dans ces logements, le questionnaire du recensement. Dans la moitié de ces IRIS (12 IRIS), les enquêteurs disposaient de l'information sur les logements voisins des logements à recenser. Dans les autres IRIS, on a réalisé un sondage d'« adresses » en effectuant un tirage par îlot. On a choisi les adresses de manière à obtenir un taux de sondage de 8 % de logements, environ, dans chaque IRIS.

En résumé, le plan d'expérience est le suivant :

* INSEE Rhône-Alpes, programme de rénovation du recensement de la population, 165, rue Garibaldi, 69401 Lyon Cedex 03.
E-mail : barbara.christian@insee.fr

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

Les secteurs confiés aux enquêteurs sont composés généralement de deux IRIS contigus : un IRIS interrogé en mode de sondage «logements» et un IRIS interrogé en mode de sondage d'«adresses»; ces deux IRIS sont, a priori, de même type. La taille des secteurs est à peu près de 2 000 logements à précensurer, pour 160 logements à interroger.

2. La non réponse dans ce test

La réponse attendue dans un premier temps est une réponse concernant le logement : on cherche à le qualifier. Autrement dit, si le logement est une résidence secondaire ou un logement vacant, le logement est un «logement répondant» puisque l'on remplit, en ce cas, une feuille de logement. Il faudrait donc disposer d'une modalité «catégorie de logement indéterminée» pour mesurer la non-réponse du logement, cas non prévu dans ce test.

Dans un second temps, si le logement est une résidence principale, on cherche à obtenir une réponse concernant le ménage. Lorsque les bulletins individuels ont pu être remplis, le ménage est considéré comme un «ménage répondant». C'est ce type de non-réponse qui va nous intéresser ici. Les cas de «non-aboutissement» d'enquête auprès d'un logement (on n'a aucun bulletin individuel) sont regroupés sous le terme «non abouti». Parmi les non aboutis, seuls ceux relatifs aux résidences principales nous intéressent pour calculer le taux de réponse

$$\text{Taux de réponse} = \frac{\text{nombre de ménages répondants}}{\text{nombre de résidences principales interrogées}}$$

3. Les résultats

On commencera par présenter les résultats des non aboutis à l'enquête avant de détailler l'analyse de la non-réponse.

3.1. Les «non aboutis» à l'enquête

	Logements	Adresses
Logement vacant	4.5	8.6
Résidence secondaire	0.3	1.2
Refus	11.9	9.2
Impossible à joindre (IAJ)	15.5	12.3
Absent de longue durée	3.0	2.9
Autre	0.3	1.2
Local à usage autre que d'habitation	0.9	2.0
Logement fusionné	0.2	0.2
Logement détruit	0.1	0.0
Logement non trouvé (plan «logements»)	0.3	0.0

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

On observe que le taux de logements vacants est plus faible au sein des secteurs sondés selon le mode « logements » alors que les résultats du recensement de 1999 ne montraient pas plus de 1 point de différence dans la vacance entre les secteurs sondés. On observe l'inverse pour le taux de ménages impossibles à joindre (ménages IAJ pour lesquels les enquêteurs n'ont pas réussi à récupérer l'information).

3.2. La non-réponse à l'enquête

Le taux de réponse globale s'élève à 69 % sur un ensemble de 3050 résidences principales :

- il est de 67 % pour les ménages interrogés selon le mode « logements » ;
- il est de 71 % pour les ménages interrogés selon le mode « adresses ».

On peut étudier la réponse à l'enquête suivant différents facteurs :

- selon le mode de sondage ;
- selon la région et le profil de l'enquêteur ;
- selon le type d'IRIS (bâti) et la taille des adresses ;
- selon les caractéristiques propres aux ménages (données non disponibles pour ces tests).

Pour hiérarchiser l'importance de chacun de ces facteurs sur la non-réponse, et neutraliser les effets structure, la modélisation de la probabilité de réponse des ménages est construite avec un modèle de type LOGIT.

3.3. La modélisation de la réponse

3.3.1. Le modèle LOGIT

On cherche à identifier les facteurs favorisant la réponse à l'enquête et on veut identifier l'effet spécifique de chaque facteur sur la probabilité de réponse. Quand le phénomène étudié est continu, la méthode appropriée est l'analyse de la variance. Dans le cas présent, le phénomène étudié est binaire : un ménage répond ou non. Un modèle de type LOGIT (basé sur la loi logistique) est adapté pour répondre à ce type d'analyse et permet de travailler « toutes choses égales par ailleurs ».

Les résultats de ce modèle s'apparentent à ceux sortis d'une régression : test de validité globale du modèle, estimation des coefficients de régression, test de significativité de chaque coefficient (test de Wald), écart-type. À l'aide de la fonction de répartition de la loi logistique, on peut également interpréter les résultats en termes de probabilité de réponse ou non réponse, conditionnellement aux variables explicatives introduites dans le modèle.

La variable d'intérêt est la variable dichotomique Y :

$$Y = \begin{cases} 1 & \text{si le ménage a répondu} \\ 0 & \text{sinon} \end{cases}$$

Réponse = variables enquêteur + type de bâti + taille de l'adresse + région + type de sondage

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

Les variables explicatives du modèle de réponse sont l'ensemble des modalités relevant des dimensions explicatives retenues ci-dessus à savoir :

Variable	Modalités
Sexe de l'enquêteur	H/F
Tranche d'âge de l'enquêteur	– moins de 30 ans / 30 ans ou plus
Source du recrutement	– ANPE – candidatures spontanées – réseau INSEE
Diplôme de l'enquêteur	– CAP, BEP et études avant bac – Bac – études supérieures au Bac
Expérience(s) professionnelle(s) antérieure(s)	– pas d'expérience professionnelle – a déjà travaillé (hors stage court études)
Type de bâti dominant de l'IRIS	individuel / ancien / concentré / hétérogène
Taille de l'adresse	– logement appartenant à une adresse < 10 logts – logement appartenant à une adresse de 10 à 30 logts – logement appartenant à une adresse > 30 logts
Région	DR1 / DR2 / DR3 (les trois régions participant au test)
Type de sondage	mode « logements » / mode « adresses »

Le modèle étant ainsi spécifié, les variables ne sont pas indépendantes puisqu'un ménage n'appartient qu'à un et un seul type d'IRIS, n'est sondé que selon un et un seul mode de sondage, etc. Pour éliminer cette dépendance « de fait », on détermine une situation de référence, et on interprète les écarts de probabilité de réponse par rapport à cette situation de référence.

La combinaison de référence a été choisie afin d'obtenir le meilleur taux de réponse possible, compte tenu des premiers résultats obtenus lors des tris croisés :

- enquêteur de sexe féminin, de plus de 30 ans, embauchée par le réseau des connaissances INSEE, de niveau Bac et ayant déjà eu une expérience professionnelle ;
- dans un IRIS « hétérogène » de la DR1, et des adresses de 10 à 30 logements ;
- avec le mode de sondage « adresses ».

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

Le passage du coefficient de régression estimé à la probabilité de réponse estimée se comprend bien une fois le modèle théorique spécifié :

$$\ln \left(\frac{p_i}{1 - p_i} \right) = A + \sum_{i=1}^q b_i X_i$$

avec p_i la probabilité de réponse

X_i les variables explicatives de la non-réponse

on en déduit :

– pour la situation de référence la probabilité de réponse p_0 :

$$p_0 = \frac{1}{1 + \exp(-A)}$$

– pour toute autre situation qui ne dévie de la situation de référence que par X_i

$$p_i = \frac{1}{1 + \exp(-A - b_i)}$$

Le test de Wald permet de juger si les coefficients sont significativement non nuls (niveau de signification : 5 %).

Il est prudent de vérifier la dispersion des probabilités individuelles de réponse.

3.3.2. Les résultats du modèle de réponse

On présente les coefficients significatifs issus de la régression (b_i) ainsi que leur écart type (Sigma b_i) et la probabilité de réponse associée à chaque facteur (taux de réponse estimé). Les paramètres sont classés du plus influent au moins influent au vue de la valeur du coefficient estimé et on a retenu uniquement les paramètres significatifs au sens du test de Wald au niveau de 5 % (ou Wald $\text{Khi}^2 > 4$).

Les probabilités individuelles estimées de réponse se situent entre 38 % et 92 % avec une moyenne à 69 % et un écart-type de 12 points, ce qui est satisfaisant. Il est également rassurant de constater que la moyenne des probabilités de réponse modélisées est proche (ici égale) au taux de réponse mesuré directement à partir de la collecte.

L'effet du mode de sondage est négligeable sur la réponse obtenue dans ce test. En revanche, les caractéristiques de l'enquêteur semblent influencer la réponse. Il est assez délicat de conclure quant au sexe des enquêteurs embauchés; il pourrait s'agir d'une disponibilité plus grande du public féminin pour assurer ce genre de mission du début à la fin. Un autre facteur mis en évidence est le fait d'avoir déjà travaillé, signe d'une meilleure organisation du travail et donc d'une plus grande efficacité à obtenir les questionnaires.

Les facteurs relatifs aux conditions de collecte ressortent dans les variables caractérisant l'habitat : si recenser une grosse adresse influe très peu à la baisse sur la probabilité de réponse, la collecte dans des quartiers à dominante d'habitat ancien présente une réelle difficulté. L'architecture de ces quartiers

INFLUENCE DU MODE DE SONDAGE SUR LA RÉPONSE DES MÉNAGES

	Paramètre b_i	Sigma (b_i)	Wald Khi ²	Taux de réponse estimé	Écart à la situation de référence
Référence	2,45	0,28	76,25	92 %	
Homme	-1,02	0,14	54,44	81 %	-11
Diplôme < Bac	-0,95	0,34	8,01	82 %	-10
Moins de 30 ans	-0,93	0,19	23,71	82 %	-10
IRIS d'habitat ancien	-0,90	0,23	15,52	83 %	-9
Diplôme > Bac	-0,54	0,17	9,94	87 %	-5
Sans expérience prof.	-0,54	0,13	17,47	87 %	-5
Adresse de + de 30 logts	-0,24	0,11	4,78	90 %	-2
Sondage « logements »	-0,18	0,08	4,95	91 %	-1
Recrutement ANPE	-0,56	0,36	2,47	ns	
Candidature spontanée	-0,30	0,30	1,01	ns	
IRIS d'habitat individuel	0,12	0,29	0,17	ns	
Adresse de - de 10 logts	0,15	0,11	1,92	ns	
IRIS d'habitat concentré	0,17	0,14	1,48	ns	
DR2	0,17	0,18	0,88	ns	
DR3	0,53	0,61	0,76	ns	

est parfois complexe et les logements vacants sont souvent nombreux. De plus, les habitants, plutôt des personnes jeunes ou, à l'opposé, âgées, sont difficiles à joindre (absentes ou refusant de répondre).

Et les effets croisés ? Ils sont très peu significatifs et ceux qui le sont ne font que renforcer les conclusions déjà mises en évidence par les effets principaux.

En conclusion, on constate que l'effet du mode de sondage (par logement ou par adresse) dans ce test est peu significatif, voire négligeable, sur la probabilité de réponse. Parmi les facteurs pris en compte dans cette analyse, l'effet relatif aux caractéristiques des enquêteurs embauchés est plus important que celui dû aux difficultés de collecte lié au type de quartier. Ces résultats ne font que confirmer des éléments déjà mis en évidence lors d'autres expériences de collecte, notamment dans les enquêtes de l'INSEE auprès des ménages.

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE : THÉORIE ET PRATIQUE

Jean DUMAIS * ¹

1. Une démarche possible pour la production des estimations détaillées

Alors que les estimations globales d'une année utilisent la collecte annuelle de l'année et les poids afférents à cette collecte pour extrapoler les résultats au niveau de la population, les estimations détaillées relatives à cette même année considèrent cette collecte comme un jeu incomplet de données (1/5), qu'il va falloir compléter (4/5) à l'aide d'un modèle. Ce modèle s'appuie sur les collectes déjà réalisées des autres groupes de rotation et entourant la collecte de l'année A, ainsi que sur de l'information auxiliaire contenue dans des fichiers administratifs, disponibles chaque année.

On peut décrire la mécanique de synthèse de la façon suivante. Il s'agit de trouver un ensemble de poids tels qu'un échantillon observé et mesuré une certaine année soit pertinent pour renseigner à propos de la population quelques années plus tard (voire plus tôt). Chaque unité statistique de base d'un groupe de rotation se voit attribuer un poids de sorte qu'elle compte pour le « bon » nombre d'unités de la population de l'année cible. Il faut donc un poids, ou un facteur d'ajustement, pour les ménages et logements d'une part, et un pour les personnes d'autre part. Ces facteurs seront obtenus par l'exploitation de sources administratives et la comparaison de quelques statistiques calculées sur l'une et l'autre sources pour la même unité statistique. Ces facteurs sont ensuite portés par chaque unité statistique de base et utilisés dans le calcul des totaux pour l'année cible. Enfin, il faut garantir qu'une statistique calculée à partir du fichier des estimations détaillées correspond au total pour la même statistique publiée deux ans auparavant à partir du fichier des estimations globales.

Ainsi, l'idée maîtresse du modèle d'imputation est d'utiliser la similarité entre données collectées et données administratives, sur quelques variables

* INSEE Rhône-Alpes, programme de rénovation du recensement de la population, 165, rue Garibaldi, 69401 Lyon Cedex 03

1. Jean Dumais a rejoint Statistique Canada et a été remplacé le 1^{er} novembre 2001 par Jean Marie Grosbras.

clés renseignées par les deux sources, pour imputer « ce qu'aurait dû être la collecte ». Il s'agit de déterminer le taux d'évolution moyen de ces variables dans les fichiers administratifs, et d'appliquer ce taux aux données collectées de l'année enquêtée afin de les « ramener » à l'année cible, puis de tout caler au total connu de cette année cible, déjà publié au titre des estimations globales.

Idéalement, on voudrait pouvoir fournir des estimations détaillées pour l'année en cours. Cependant, le temps nécessaire à la réception et au traitement préalable (géocodage, apurement, appariement, etc.) des fichiers administratifs et le temps nécessaire à la préparation des modèles imposent un délai de deux ans ; par exemple la taxe d'habitation (TH) au 1^{er} janvier 2000 reçue à l'INSEE sera prête à utiliser à partir de juin 2001 ; en mode réel, il est possible que la TH de l'année A ne soit pas prête au moment où il faudra calculer les paramètres du modèle pour A-2. Finalement, il faut aussi compter qu'il faut traiter une dizaine de paramètres pour quelque 15 000 IRIS et 22 régions, soit environ 150 000 modèles.

Cette mise en garde faite, imaginons que les fichiers administratifs soient instantanément disponibles, en A pour A. On pourrait alors envisager de produire en A les données détaillées pour A en n'utilisant que les cinq collectes les plus récentes. Nombre de commentateurs ont fait valoir leurs réticences à voir des données vieillies sur cinq ans sans autre calage temporel, source de dérapage éventuel. Nous retiendrons ici le calcul en A de données détaillées pour A-2.

2. Éléments théoriques

Pour simplifier la présentation, on suppose que le problème décrit ici concerne les petites communes d'une région donnée. La solution proposée vaut pour toute région. La solution est immédiatement transposable en grande commune en remplaçant « petite commune » par « adresse » et « région » par « zone RRP ».

Examinons le tableau ci-dessous :

- nous sommes en A et nous voulons estimer les résultats relatifs à A-2 ;
- nous disposons au total de 7 années de collecte (R) : 4 sont antérieures à l'année de référence (groupes I, II, IV et V), deux lui sont postérieures (groupes IV et V) et une est datée A-2 (groupe III) ;
- nous disposons des fichiers administratifs (F), chaque année, au niveau « fin » pour toute la population de la région d'intérêt.

Les données des communes non interrogées en A-2 donnent lieu à une imputation par modèle comme il est décrit dans les paragraphes qui suivent.

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

G	A-6		A-5		A-4		A-3		A-2		A-1		A	
I		F		F	R	F		F	?	F		F		F
II		F		F		F	R	F	?	F		F		F
III		F		F		F		F	R	F		F		F
IV	R	F		F		F		F	?	F	R	F		F
V		F	R	F		F		F	?	F		F	R	F

Ainsi, pour le groupe I, on appliquerait le rapport moyen F_{A-2}/F_{A-4} calculé sur l'ensemble des communes ayant participé en A-4 aux valeurs de la collecte relatives à A-4 (R_{A-4}) pour imputer une valeur de « recensement » à chaque commune du groupe I en A-2. On voit que pour les groupes IV et V, on peut calculer ces rapports à partir de collectes antérieures (respectivement A-6 et A-5) et postérieures (respectivement A-1 et A). Pour ces deux groupes, il existe deux sources d'imputation qu'il est possible de combiner.

Posons $\tilde{Y}_j^i$ la valeur imputée pour la variable d'intérêt Y pour le groupe de rotation i , $i = I, \dots, V$, à partir des données collectées en l'année j . Le tableau ci-dessus peut maintenant être complété :

G	A-6		A-5		A-4		A-3		A-2		A-1		A	
I		F		F	R→	F	→	F	$\tilde{Y}_{A-4}^I$	F		F		F
II		F		F		F	R→	F	$\tilde{Y}_{A-3}^{II}$	F		F		F
III		F		F		F		F	R_{A-2}	F		F		F
IV	R→	F	→	F	→	F	→	F	$\tilde{Y}_{A-6/A-1}^{IV}$	F	← R	F		F
V		F	R→	F	→	F	→	F	$\tilde{Y}_{A-5/A}^V$	F	←	F	← R	F

L'estimation détaillée pour une variable Y donnée en A-2 s'écrit comme suit :

$$\tilde{\tilde{Y}}_{A-2} = \underbrace{\tilde{Y}_{A-4}^I + \tilde{Y}_{A-3}^{II}}_{\text{extrapolations}} + \underbrace{R_{A-2}}_{\text{observations}} + \underbrace{\tilde{Y}_{A-6/A-1}^{IV} + \tilde{Y}_{A-5/A}^V}_{\text{extra-rétropolations}}$$

Examinons une partie du modèle, par exemple l'imputation du Groupe II donnant lieu à $\tilde{Y}_{A-3}^{II}$:

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

Groupe	Commune	A-3			A-2		A-1	
II	c	R _{c, A-3}	F _{c, A-3}	→	$\tilde{Y}_{c, A-3}^I$	F _{c, A-2}		F _{c, A-1}
	d	R _{d, A-3}	F _{d, A-3}	→	$Y_{d, A-3}^I$	F _{d, A-2}		F _{d, A-1}
III	p		F _{p, A-3}		R _{p, A-2}	F _{p, A-2}		F _{p, A-1}
	q		F _{q, A-3}		R _{q, A-2}	F _{q, A-2}		F _{q, A-1}

Sur l'ensemble des communes ayant été sollicitées en A-3, la valeur moyenne dans les fichiers administratifs est

$$\overline{F}_{A-3} = \frac{F_{c, A-3} + F_{d, A-3}}{2}$$

et la valeur moyenne dans les fichiers administratifs sur l'ensemble des communes sollicitées en A-2 est

$$\overline{F}_{A-2} = \frac{F_{c, A-2} + F_{d, A-2}}{2}$$

On cherche à évaluer $\tilde{Y}_{A-3}^{II}$, la valeur obtenue à partir de A-3, pour chaque commune du groupe II en fonction de l'évolution de l'information auxiliaire contenue dans les fichiers, autrement dit :

$$\tilde{Y}_{k, A-3}^{II} = R_{k, A-3} \times \frac{\overline{F}_{A-2}}{\overline{F}_{A-3}} = R_{k, A-3} \times \theta_{A-3} ,$$

avec $\theta_{A-3} = \frac{\overline{F}_{A-2}}{\overline{F}_{A-3}}$ et $k \in$ Groupe II.

La technique est la même pour $\tilde{Y}_{A-4}^I$, l'information obtenue pour les communes du groupe I à partir de l'année A-4. Elle est un peu plus lourde pour les imputations correspondant aux «extra-rétropolations». Dans ces extra-rétropolations, on combine l'information tirée d'une source amont et celle tirée d'une source aval (groupes IV et V), et on privilégie la collecte la plus proche de l'année cible. Détaillons par exemple $\tilde{Y}_{A-6/A-1}^{IV}$:

$$\tilde{Y}_{A-6/A-1}^{IV} = 0,2 \times R_{A-6} \times \theta_{A-6} + 0,8 \times R_{A-1} \times \theta_{A-1}$$

où les coefficients θ_{A-6} et θ_{A-1} sont obtenus de manière équivalente à ce qui est décrit pour A-3. Pour le groupe V, on aura

$$\tilde{Y}_{A-5/A}^V = 0,4 \times R_{A-5} \times \theta_{A-5} + 0,6 \times R_A \times \theta_A.$$

Une fois tous les coefficients θ obtenus, il faut s'assurer que les estimations qu'on en dériverait correspondent aux totaux déjà publiés. Pour toute variable Y , les données collectées dans les communes recensées en A-2 avaient été extrapolées à l'ensemble de l'univers des communes. L'estimation pour le total

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

des communes de la variable y , appelée estimation globale, avait alors été :

$$\hat{Y}_{A-2} = \sum w_l R_{l,A-2}$$

où les poids w_l proviennent du plan de sondage exécuté en A-2.

Deux ans plus tard, en A, on reconstitue par synthèse les données pour les groupes non observés en A-2. On peut calculer alors une autre estimation pour A-2, appelée estimation détaillée :

$$\tilde{\tilde{Y}}_{A-2} = \underbrace{\tilde{Y}_{A-4}^I + \tilde{Y}_{A-3}^{II}}_{\text{extrapolations}} + \underbrace{R_{A-2}}_{\text{observations}} + \underbrace{\tilde{Y}_{A-6/A-1}^{IV} + \tilde{Y}_{A-5/A}^V}_{\text{extra-rétropolations}};$$

il faut maintenant faire en sorte que

$$\tilde{\tilde{Y}}_{A-2} = \hat{Y}_{A-2}.$$

Cette égalité est facilement réalisée en corrigeant les totaux estimés selon les catégories de population à partir desquels les coefficients θ ont été calculés; on transforme les coefficients θ en coefficients calés aux estimations globales θ^* de la façon suivante

$$\theta^* = \frac{\hat{Y}_{A-2} - R_{A-2}}{\tilde{\tilde{Y}}_{A-2} - R_{A-2}} \times \theta.$$

Ces coefficients calés seront archivés et utilisés pour le calcul des tableaux détaillés.

3. Un exemple

On étudie l'exemple d'une région de 10 communes, sondée par rotation de panels de taille 2. Le sondage est en régime permanent, de sorte que toutes les collectes et tous les fichiers administratifs nécessaires sont disponibles. Le tableau ci-dessous donne le résultat de la collecte menée en A-2 et les comptes administratifs pour quatre variables d'intérêt (nombre de femmes et nombre d'hommes pour les tranches d'âge «0 à 39 ans» et «40 ans ou plus») dans chacune des deux communes du groupe de rotation III :

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

Groupe	Commune	Variable	Recensement	Administratif
III	6	F00-39	4	4
III	6	F40+	20	22
III	6	H00-39	15	20
III	6	H40+	16	20
III	7	F00-39	17	20
III	7	F40+	18	20
III	7	H00-39	15	20
III	7	H40+	18	20
III	Toutes	F00-39	21	24
Tous	Toutes	F00-39	105	120
III	Toutes	Total	123	146
Tous	Toutes	Total	615	730

À partir des résultats du recensement tenu en A-2, on peut estimer la population totale à 615 personnes pour l'année A-2, résultat global pour la région. On dispose aussi de données administratives à propos de ces quatre variables pour toutes les communes depuis A-6, et jusqu'à A.

Pour le groupe II (voir plus bas), le coefficient θ_{F0039} est donné par

$$\theta_{F0039} = \frac{\frac{(13 + 50)}{2}}{\frac{(12 + 55)}{2}} = 0,94.$$

Les autres coefficients sont calculés de même ce qui permet de remplir la colonne « Recensement » pour A-2. Finalement, une fois la colonne remplie, on estime la population totale à $\tilde{Y}_{A-2} = 595,40$.

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

Groupe	Commune	Variable	A 3			A 2		
			Recensement / Imputation	Administratif	θ_{A3}	Recensement / Imputation	Administratif	θ_{A2}
II	3	F00-39	12	12	0,94	12x0,94	13	
II	3	F40+	13	11	0,92	13x0,92	11	
II	3	H00-39	14	13	0,93	14x0,93	13	
II	3	H40+	30	28	0,90	30x0,90	25	
II	4	F00-39	60	55	0,94	60x0,94	50	
II	4	F40+	25	26	0,92	25x0,92	23	
II	4	H00-39	18	16	0,93	18x0,93	14	
II	4	H40+	12	11	0,90	12x0,90	10	
III	6	F00-39		4		4	4	1
III	6	F40+		24		20	22	1
III	6	H00-39		20		15	20	1
III	6	H40+		20		16	20	1
III	7	F00-39		20		17	20	1
III	7	F40+		20		18	20	1
III	7	H00-39		20		15	20	1
III	7	H40+		20		18	20	1
III	Toutes	F00-39	observation			$\Sigma 1 = 21$		
Tous	Toutes	F00-39	observation + synthèse			$\Sigma \theta = 146,12$		
III	Toutes	total	observation			$\Sigma 1 = 123$		
Tous	Toutes	total	observation + synthèse			$\Sigma \theta = 595,40$		

La dernière ligne du tableau nous indique, qu'après synthèse, on estimerait la population de la région à 596 personnes ; or, les estimations globales publiées en A-2 établissaient la taille de cette population à 615 personnes.

Il est nécessaire de caler les synthèses aux estimations globales pour que les deux séries coïncident. Par ailleurs, un examen rapide du tableau montre que tous les groupes d'âge-sexe n'ont pas évolué de façon uniforme. La population des femmes de moins de 40 ans est estimée à 146,12 personnes (estimation détaillée) alors que l'estimation globale correspondante était de 105 personnes. Pour les communes hors du groupe III, on doit donc ajuster le total estimé par un facteur $\frac{105 - 21}{146,12 - 21} = \frac{84}{125,12} = 0,6714$ pour que les estimations détaillées correspondent aux estimations globales.

Les synthèses se caleront donc sur ces groupements âge-sexe. Ainsi, on peut calculer le tableau des ajustements nécessaires en estimant la population de chaque groupe d'âge-sexe pour les communes non enquêtées en A-2, soit par le modèle de synthèse en y faisant la somme des valeurs pertinentes, soit en utilisant le plan de sondage. On obtient alors les résultats montrés ci-dessous (les tableaux de données après collecte, après imputation et après calage sont donnés à la fin du texte) :

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

	Population de la région		Population hors communes 6 et 7		Coefficient de calage
	d'après la synthèse	d'après le plan de sondage	d'après la synthèse	d'après le plan de sondage	
F00-39	146,12	105	125,12	84	0,6714
F40+	140,49	190	102,49	152	1,4831
H0039	131,20	150	101,20	120	1,1858
H40+	177,60	170	143,60	136	0,9471
Total	595,40	615	472,40	492	

Enfin, on peut calculer l'ensemble des populations ajustées, comme montré dans ce tableau-ci :

Groupe	Commune	Estimations détaillées pour A-2			
		Recensement	Administratif	Imputation	imputation calée
I	1		48	38,34	41,93
I	2		48	46,93	51,52
II	3		62	63,19	66,25
II	4		97	106,92	102,02
III	6		66	55,00	55,00
III	7	68	80	68,00	68,00
IV	9		17	23,51	24,32
IV	10		24	31,49	34,82
V	11		84	72,23	73,40
V	12		88	89,80	97,75
Total	Total	615	614	595,40	615,00

4. Observations finales

Comme on a pu le voir dans cette note, la mécanique de synthèse est relativement simple, mais elle doit être mise en œuvre avec soin étant donné le nombre d'opérations et le nombre de structures dont il faut tenir compte. Une condition nécessaire à la qualité de l'ajustement des données « détaillées » est la qualité des sources administratives ; sans elle, aucun modèle de calage ne saurait donner de résultats satisfaisants. Notre prochain travail sur ce sujet est de monter un exemple « vrai » de fichier détail pour quelque partie des grandes communes de Rhône-Alpes, en utilisant la TH 1999 et 2000 et les résultats de l'exploitation exhaustive du RP99. Cet exemple ne pourra raisonnablement porter que sur les variables « ménage » et « logement ». Ultérieurement, avec l'accès à une source administrative concernant les personnes (FCAM, IRPP ou autre), il sera possible de monter un exemple de fichier détail pour les variables « personnes ».

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

Annexe : Tableaux détaillés

TABLEAU 1. – Données après collecte

Groupe		A-6			A-5			A-4			A-3			A-2			A-1			A		
Commune	Variable	Rec	Adm	0	Rec	Adm	0	Rec	Adm	0	Rec	Adm	0	Rec	Adm	0	Rec	Adm	0	Rec	Adm	0
I 1	F00-39	11			9			11	10	0,81	11			13			12			13		
1	F40+	13			11			11	12	0,96	12			11			12			13		
1	H00-39	17			14			11	15	0,93	13			13			12			12		
1	H40+	18			16			13	18	0,67	14			11			12			12		
I 2	F00-39	24			20			13	22	0,81	15			13			12			11		
2	F40+	12			10			13	11	0,96	15			11			12			11		
2	H00-39	14			12			15	13	0,93	14			13			12			10		
2	H40+	17			14			15	15	0,67	13			11			12			10		
II 3	F00-39	17			15			17			12	12	0,94	13			12			9		
3	F40+	20			17			19			13	11	0,92	11			12			9		
3	H00-39	17			14			15			14	13	0,93	13			12			11		
3	H40+	29			25			27			30	28	0,90	25			24			28		
II 4	F00-39	58			48			52			60	55	0,94	50			47			54		
4	F40+	25			23			25			25	26	0,92	23			22			25		
4	H00-39	15			14			15			18	16	0,93	14			14			16		
4	H40+	10			9			10			12	11	0,90	10			9			10		
III 6	F00-39	5			4			4			4			4	4	1,00	3			3		
6	F40+	25			21			23			24			20	22	1,00	20			20		
6	H00-39	20			18			20			20			15	20	1,00	20			20		
6	H40+	23			19			20			20			16	20	1,00	20			20		
III 7	F00-39	20			20			20			20			17	20	1,00	20			20		
7	F40+	20			20			20			20			18	20	1,00	20			20		
7	H00-39	20			20			20			20			16	20	1,00	20			20		
7	H40+	20			20			20			20			18	20	1,00	20			20		
IV 9	F00-39	5	6	0,93	5			5			5			5			5	5	1,08	5		
9	F40+	3	2	1,00	2			2			1			2			3	1	1,20	2		
9	H00-39	5	5	0,86	4			4			3			4			5	3	1,50	4		
9	H40+	5	6	0,91	5			6			5			6			7	5	1,25	6		
IV 10	F00-39	9	8	0,93	7			8			7			8			9	7	1,08	8		
10	F40+	9	10	1,00	9			10			9			10			11	9	1,20	10		
10	H00-39	3	2	0,86	2			2			1			2			3	1	1,50	2		
10	H40+	3	5	0,91	4			4			3			4			5	3	1,25	4		
V 11	F00-39	7			6	6	1,15	6			5			6			5			7	6	0,92
11	F40+	9			7	8	1,15	8			7			8			7			9	8	0,94
11	H00-39	11			8	9	1,16	10			9			10			9			11	10	0,95
11	H40+	63			50	56	1,11	60			50			60			50			43	60	0,98
V 12	F00-39	16			12	14	1,15	15			16			17			18			20	19	0,92
12	F40+	21			20	18	1,15	20			21			22			23			20	24	0,94
12	H00-39	27			20	23	1,16	25			26			27			28			30	29	0,95
12	H40+	20			20	18	1,11	20			21			22			23			30	24	0,98

615 614

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

TABLEAU 2. – Après imputation de A-2

Commune	Variable	A-6		A-5		A-4		A-3		A-2		A-1		A	
		Rec	Adm 0	Rec	Adm 0	Rec	Adm 0	Rec	Adm 0	Rec	Adm 0	Rec	Adm 0	Rec	Adm 0
I 1	F00-39	11		9		11	10 0,81	11		8,94	13	12		13	
1	F40+	13		11		11	12 0,96	12		10,52	11	12		13	
1	H00-39	17		14		11	15 0,93	13		10,21	13	12		12	
1	H40+	18		16		13	18 0,67	14		8,67	11	12		12	
I 2	F00-39	24		20		13	22 0,81	15		10,56	13	12		11	
2	F40+	12		10		13	11 0,96	15		12,43	11	12		11	
2	H00-39	14		12		15	13 0,93	14		13,93	13	12		10	
2	H40+	17		14		15	15 0,67	13		10,00	11	12		10	
II 3	F00-39	17		15		17		12	12 0,94	11,28	13	12		9	
3	F40+	20		17		19		13	11 0,92	11,95	11	12		9	
3	H00-39	17		14		15		14	13 0,93	13,03	13	12		11	
3	H40+	29		25		27		30	28 0,90	26,92	25	24		28	
II 4	F00-39	58		48		52		60	55 0,94	56,42	50	47		54	
4	F40+	25		23		25		25	26 0,92	22,97	23	22		25	
4	H00-39	15		14		15		18	16 0,93	16,76	14	14		16	
4	H40+	10		9		10		12	11 0,90	10,77	10	9		10	
III 6	F00-39	5		4		4		4		4,00	41,00	3		3	
6	F40+	25		21		23		24		20,00	221,00	20		20	
6	H00-39	20		18		20		20		15,00	201,00	20		20	
6	H40+	23		19		20		20		16,00	201,00	20		20	
III 7	F00-39	20		20		20		20		12,00	201,00	20		20	
7	F40+	20		20		20		20		13,00	201,00	20		20	
7	H00-39	20		20		20		20		15,00	201,00	20		20	
7	H40+	20		20		20		20		18,00	201,00	20		20	
IV 9	F00-39	5 6 0,93		5		5		5		5,26	5	5 5 1,08		5	
9	F40+	3 2 1,00		2		2		1		3,48	2	3 1 1,20		2	
9	H00-39	5 5 0,86		4		4		3		6,86	4	5 3 1,50		4	
9	H40+	5 6 0,91		5		6		5		7,91	6	7 5 1,25		6	
IV 10	F00-39	9 8 0,93		7		8		7		9,47	8	9 7 1,08		8	
10	F40+	9 10 1,00		9		10		9		12,36	10	11 9 1,20		10	
10	H00-39	3 2 0,86		2		2		1		4,11	2	3 1 1,50		2	
10	H40+	3 5 0,91		4		4		3		5,55	4	5 3 1,25		4	
V 11	F00-39	7		6 6 1,15		6		5		6,62	6	5		7 6 0,92	
11	F40+	9		7 8 1,15		8		7		8,29	8	7		9 8 0,94	
11	H00-39	11		8 9 1,16		10		9		9,96	10	9		11 10 0,95	
11	H40+	63		50 56 1,11		60		50		47,35	60	50		43 60 0,98	
V 12	F00-39	16		12 14 1,15		15		16		16,56	17	18		20 19 0,92	
12	F40+	21		20 18 1,15		20		21		20,48	22	23		20 24 0,94	
12	H00-39	27		20 23 1,16		25		26		26,33	27	28		30 29 0,95	
12	H40+	20		20 18 1,11		20		21		26,44	22	23		30 24 0,98	
595,40 614															

LE FONCTIONNEMENT DE L'ESTIMATION DÉTAILLÉE

TABLEAU 3. – Estimations ajustées après calage pour les enregistrements « personne »

Au final, pour l'année A 2							
Groupe	Commune	Variable	Recensement	Administratif	Imputation	Calage au Global	Imputation calée
I	1	F00-39		13	8,938	0,6714	6,00
I	1	F40+		11	10,522	1,4831	15,60
I	1	H00-39		13	10,214	1,1858	12,11
I	1	H40+		11	8,667	0,9471	8,21
I	2	F00-39		13	10,563	0,6714	7,09
I	2	F40+		11	12,435	1,4831	18,44
I	2	H00-39		13	13,929	1,1858	16,52
I	2	H40+		11	10,000	0,9471	9,47
II	3	F00-39		13	11,284	0,6714	7,58
II	3	F40+		11	11,946	1,4831	17,72
II	3	H00-39		13	13,034	1,1858	15,46
II	3	H40+		25	26,923	0,9471	25,50
II	4	F00-39		50	56,418	0,6714	37,88
II	4	F40+		23	22,973	1,4831	34,07
II	4	H00-39		14	16,759	1,1858	19,87
II	4	H40+		10	10,769	0,9471	10,20
III	6	F00-39	20	4	4		20
III	6	F40+	20	22	20		20
III	6	H00-39	20	20	13		20
III	6	H40+	20	20	16		20
III	7	F00 39	17	20	17		17
III	7	F40+	18	20	18		18
III	7	H00-39	15	20	15		15
III	7	H40+	18	20	18		18
IV	9	F00-39		5	5,262	0,6714	3,53
IV	9	F40+		2	3,480	1,4831	5,16
IV	9	H00-39		4	6,857	1,1858	8,13
IV	9	H40+		6	7,909	0,9471	7,49
IV	10	F00-39		8	9,471	0,6714	6,36
IV	10	F40+		10	12,360	1,4831	18,33
IV	10	H00-39		2	4,114	1,1858	4,88
IV	10	H40+		4	5,545	0,9471	5,25
V	11	F00 39		6	6,624	0,6714	4,45
V	11	F40+		8	8,293	1,4831	12,30
V	11	H00-39		10	9,962	1,1858	11,81
V	11	H40+		60	47,348	0,9471	44,84
V	12	F00-39		17	16,560	0,6714	11,12
V	12	F40+		22	20,481	1,4831	30,37
V	12	H00 39		27	26,327	1,1858	31,22
V	12	H40+		22	26,436	0,9471	25,04
Total	Total	Total	123	614	595,402		615,00

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS AVEC LE RECENSEMENT RÉNOVÉ

Brigitte BACCAÏNI *

1. Le RRP et la question des migrations

1.1. Quelques rappels sur les enjeux de la question sur les migrations

Depuis 1962, on demande aux individus recensés d'indiquer leur résidence au 1^{er} janvier de l'année du recensement précédent. C'est à partir de l'exploitation de cette question que sont publiées les principales statistiques sur les migrations internes en France. Pour chaque zone du territoire sont également publiés les « soldes migratoires apparents » : différence entre la variation de la population et le solde naturel.

Rappel des problèmes et limites liées à la formulation de la question :

- On dispose d'une information sur des *migrants* (des individus dont le lieu de résidence à la date du recensement est différent de celui qu'ils occupaient lors du précédent recensement) et non sur des *migrations*. Du fait des migrations multiples et des retours (d'autant plus nombreux que la période intercensitaire est longue), ce nombre de migrants est toujours inférieur au nombre réel de migrations.

Ce problème est plus gênant lors d'analyses de la « mobilité » (fréquence des changements de logement, de commune, etc) que pour l'analyse des « flux entre zones ». En effet, pour une connaissance du rôle des migrations dans le processus de redistribution spatiale de la population, l'information sur le nombre de migrants entre zones, entre deux dates données, est probablement plus utile qu'une information sur les migrations.

Il faut aussi souligner que les soldes migratoires internes sont, eux, bien mesurés avec la question sur les migrants (en négligeant l'impact de la mortalité).

* (INSEE/INED), INSEE Rhône-Alpes, programme de rénovation du recensement de la population, 165, rue Garibaldi, 69401 Lyon Cedex 03 E-mail . brigitte.baccaini@insee.fr

En résumé :

- Au niveau des zones géographiques : on connaît correctement le solde migratoire interne, le nombre d'immigrants (internes et externes) et le nombre d'émigrants (internes). Mais on ne connaît pas le nombre total d'entrées et de sorties qui se sont produites dans la zone au cours de la période (ou sur une année).
- Au niveau des individus : on ne peut que comparer le lieu de résidence actuel à celui du 1^{er} janvier de l'année du précédent recensement. On ne peut pas étudier les trajectoires résidentielles précises des migrants (on ne connaît pas les éventuelles étapes intermédiaires).
- Les caractéristiques individuelles et celles du logement ne sont connues qu'à la date du recensement. En particulier, on ne connaît pas l'âge exact auquel s'est produit le mouvement : pour le RP99, les migrants âgés de 25 ans au RP ont pu effectuer leur(s) migration(s) entre 16 et 25 ans.

1.2. Les propositions faites dans le groupe de travail du CNIS

- La méthode préconisée pour le recensement rénové oblige à reconsidérer complètement la question sur le lieu de résidence antérieur.

En 1999, lors des travaux du groupe de travail « Dynamique des territoires », il avait été proposé (suite aux suggestions faites dans le rapport rédigé par B. Baccaïni ¹) de la remplacer par deux questions : une sur le lieu de résidence cinq ans auparavant, et une sur le lieu de résidence un an auparavant. En effet :

- la question sur un an aboutit à une meilleure connaissance de la mobilité résidentielle et de son évolution, et à une meilleure connaissance des trajectoires résidentielles ;
- la question sur cinq ans est préférable pour l'étude des flux entre petites zones ou entre zones distantes (ces flux étant de taille réduite).

Donc : une question sur un an permet des analyses fines de la mobilité (propension à changer de résidence) et de son évolution, à condition de travailler sur des zones de taille assez importante, ou sur un type de mobilité assez fréquent (changement de logement et non de région, par exemple).

Une question sur cinq ans est préférable, d'une part pour l'analyse de la mobilité sur des zones de petite taille, d'autre part pour analyser les flux entre zones distantes.

Pour les problématiques « logement », la question sur un an semble préférable, alors que pour les problématiques « territoires », la question sur cinq ans est sans doute préférable.

En 2000, le groupe questionnaire a considéré que la question à cinq ans était prioritaire par rapport à la question à un an.

- Il a également été proposé d'introduire des questions sur la description du logement antérieur (un an auparavant) : statut d'occupation, type d'immeuble,

1. Baccaïni Brigitte (1999), « Recensement traditionnel et recensement rénové pour l'étude des migrations résidentielles ». INSEE, 101 p

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

nombre de pièces. La Direction Générale des Collectivités Locales (DGCL) s'est montrée défavorable à l'introduction de questions de ce type dans un recensement. Par ailleurs, les résultats des tests réalisés en 2000 et 2001 ont montré que ces questions étaient trop complexes.

1.3. Les propositions du programme RRP

Questions maintenues ou adaptées

- Une question sur le lieu de résidence 5 ans auparavant serait introduite, en remplacement de la question sur le lieu de résidence au 1^{er} janvier de l'année du précédent recensement.
- La question de la feuille de logement sur l'année d'arrivée dans le logement serait conservée.

Questions demandées par les utilisateurs que le programme RRP n'envisage pas d'introduire

- La question sur le lieu de résidence un an auparavant
- Les caractéristiques du logement antérieur

Pour la suite, on part donc du principe suivant : l'analyse de la mobilité et des flux entre territoires serait désormais réalisée à partir d'une question sur le lieu de résidence 5 ans auparavant.

2. L'analyse des migrations avec le RRP

2.1. Présentation du fichier détail produit par le RRP

On suppose que le territoire est divisé en deux régions. Chaque région est composée de 10 communes. Pour constituer le fichier détail, on a appliqué les taux de mobilité par âge du RP99. On est en A et on a des estimations de population pour A-2.

groupe	A-6	A-5	A-4	A-3	A-2	A-1	A
I			R		Imputation		
II				R	Imputation		
III					R		
IV	R				Imputation	R	
V		R			Imputation		R

Les communes du groupe I : recensées en A-4

Les communes du groupe II : recensées en A-3

Les communes du groupe III : recensées en A-2

Les communes du groupe IV : recensées en A-6 et en A-1

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

Les communes du groupe V : recensées en A-5 et en A

Le fichier détail fournit, pour chaque individu :

- Des informations de localisation géographique : commune, région de résidence, groupe de la commune.
- Des informations socio-démographiques : âge, sexe, activité, statut d'occupation du logement (et toutes les autres qui seront recueillies lors du recensement).
- Le lieu de résidence antérieur (5 ans plus tôt) : même commune, autre commune même région, autre région (cette variable synthétique étant établie à partir de la comparaison de la commune de résidence actuelle et de la commune de résidence cinq ans plus tôt).
- Un «ajustement», c'est-à-dire une pondération affectée à chaque individu, qui résulte de l'imputation.

Il s'agit donc d'un fichier identique à ce que l'on a avec les recensements traditionnels, avec pour seule différence une pondération affectée à chaque individu.

2.2. Les arrivants dans chaque commune

Le tableau 1 présente le cas de la commune 1.

TABLEAU 1. – Commune 1 (groupe I) : lieu de résidence de la population en A-7

caractéristiques en A-2	même commune		autre commune même région		autre région		ensemble	
	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)
ensemble	29,6	70,6	4,7	11,2	7,6	18,2	41,9	100,0
Age								
moins de 40 ans	11,5	56,3	3,6	17,8	5,3	25,9	20,3	100,0
plus de 40 ans	18,1	84,0	1,1	5,1	2,4	10,9	21,6	100,0
Sexe								
hommes	17,1	79,3	2,5	11,6	2,0	9,1	21,6	100,0
femmes	12,5	61,3	2,2	10,8	5,7	27,9	20,3	100,0
Statut occup logt								
propriétaire	1,1	50,0	0,5	25,0	0,5	25,0	2,2	100,0
locataire	28,5	71,7	4,2	10,5	7,1	17,8	39,7	100,0
Activité								
actif occupé	10,0	56,3	4,7	26,4	3,1	17,2	17,8	100,0
chômeur	0,0		0,0		0,0		0,0	-
retraité	11,5	91,3	0,0	0,0	1,1	8,7	12,6	100,0
étudiant	4,3	60,0	0,0	0,0	2,8	40,0	7,1	100,0
autre	3,8	85,7	0,0	0,0	0,6	14,3	4,4	100,0

Source Insee, RRP, A 2

La commune 1 fait partie du groupe I. Elle a été recensée en A-4. Les estimations ont été calculées selon la méthode présentée par Jean Dumais

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

plus haut. Aux observations de A-4 on a donc appliqué un ajustement égal à : (F_{A-2}/F_{A-4}) calculé sur l'ensemble des communes du groupe I (F étant les données des fichiers administratifs). La variable « lieu de résidence 5 ans plus tôt » est traitée exactement comme les autres variables du fichier détail : on applique à chaque individu une pondération correspondant à cet ajustement. Ainsi, dans ces communes du groupe 1 : pour les hommes de moins de 40 ans, l'ajustement est de 0,545 ; pour les femmes de moins de 40 ans, il est de 1,101 ; pour les hommes de plus de 40 ans il est de 1,419 ; pour les femmes de plus de 40 ans il est de 0,631.

Si l'on veut estimer le nombre de personnes n'ayant pas changé de commune pendant les 5 années, on va appliquer à chacune des personnes ayant déclaré en A-4, ne pas avoir changé de commune, ce facteur d'ajustement. Ainsi, alors qu'on avait recensé en A-4, 33 individus dans ce cas, on n'en aurait plus que 30 en A-2.

Les tableaux 2a, 2b, 2c présentent les résultats pour les 10 communes de la région 1 :

- Les stables (individus n'ayant pas changé de commune)
- Les migrants intrarégionaux (individus ayant changé de commune dans la région)
- Les migrants interrégionaux (individus ayant changé de région)

On retrouve les taux de migration par âge introduits dans le fichier de départ, ce qui montre la robustesse du modèle.

2.3. Les échanges entre les deux régions

On suppose dans notre exemple que le territoire est composé de deux régions. On peut donc avoir une estimation des entrants, des sortants (ce sont les entrants dans l'autre région) et donc du solde migratoire de chaque région.

Dans le cas d'un territoire constitué d'un plus grand nombre de régions, le calcul des sortants d'une région donnée serait un peu plus long (il faudrait considérer les entrants dans toutes les régions en provenance de cette région, comme l'on fait avec le recensement traditionnel).

2.4. Les migrants externes

Le passage au RRP ne change rien au problème de l'estimation d'un solde externe, pour une zone quelconque (solde des entrées et des sorties avec l'étranger). Cette estimation posait déjà en effet de sérieux problèmes avec le RP traditionnel : voir Baccaïni (1999).

Théoriquement, on devrait pouvoir, pour toute zone du territoire, estimer les sorties vers l'étranger. En effet, la variation totale de la population d'une zone entre deux recensements peut être décomposée en plusieurs termes : le mouvement naturel (différence entre les naissances et les décès), le solde migratoire avec le reste du territoire métropolitain et le solde migratoire avec l'étranger ou les Dom-Tom.

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

TABLEAU 2a. – Les stables : population résidant en A-7 dans la même commune qu'en A-2 (% de la population de A-2)

caractéristiques en A-2	Communes										
	1	2	3	4	6	7	9	10	11	12	
ensemble	70,6	72,9	61,5	64,8	69,1	70,6	62,6	76,5	76,8	66,8	
Age											
moins de 40 ans	62,1	56,2	40,4	58,9	52,6	62,5	49,6	46,8	64,9	46,9	
plus de 40ans	79,5	89,4	73,7	80,4	77,8	77,8	72,6	100,0	79,9	84,4	
Sexe											
hommes	78,0	69,8	50,2	66,3	70,8	80,0	60,4	76,0	68,3	65,7	
femmes	62,8	75,6	67,8	60,0	67,7	60,6	63,9	78,1	78,8	67,7	
Statut occup logt											
propriétaire	39,9	73,4	57,3	71,6	69,1	61,3	68,8	78,6	79,6	69,6	
locataire	72,7	71,1	65,8	58,1		78,4	38,7	70,4	74,0	63,7	
Activité											
actif occupé	61,3	64,5	64,1	56,1	70,8	72,4	60,7	76,7	81,0	62,0	
chômeur		33,3	72,2	84,6	100,0	100,0	0,0	100,0	88,9	45,2	
retraité	92,0	100,0	71,3	82,2	68,8	76,5	100,0	100,0	75,4	86,8	
étudiant	61,5	70,6	40,0	61,5	42,9	56,3	56,2	33,3	80,2	45,9	
autre	80,8	83,1	39,7	67,7	80,0	75,0	17,2	82,6	46,7	90,1	

Source : Insee, RRP, A 2

TABLEAU 2b. – Les migrants intrarégionaux (entre A-7 et A-2). Population arrivée dans la commune en provenance d'une autre commune de la région (% de la population de A-2)

caractéristiques en A-2	Communes										
	1	2	3	4	6	7	9	10	11	12	
ensemble	11,2	10,4	18,1	24,5	21,8	26,5	27,6	14,3	14,2	22,6	
Age											
moins de 40 ans	16,7	16,6	25,3	29,6	26,3	31,3	36,9	32,4	22,5	38,4	
plus de 40ans	5,4	4,2	13,9	11,2	19,4	22,2	20,6	0,0	12,0	8,6	
Sexe											
hommes	11,7	13,0	25,7	25,9	25,0	17,1	39,6	15,3	28,7	18,0	
femmes	10,7	8,0	13,8	20,4	19,4	36,4	21,0	10,9	10,8	25,9	
Statut occup logt											
propriétaire	19,9	12,0	18,1	19,2	21,8	38,7	18,9	12,4	14,0	18,1	
locataire	10,6	4,8	18,1	29,9	-	16,2	61,3	19,7	14,4	27,7	
Activité											
actif occupé	24,2	14,1	12,4	33,2	16,7	24,1	21,4	13,5	12,0	25,9	
chômeur		0,0	27,8	15,4	0,0	0,0	100,0	0,0	0,0	27,3	
retraité	0,0	0,0	11,5	4,2	25,0	23,5	0,0	0,0	17,7	11,3	
étudiant	0,0	17,6	51,1	26,9	42,9	37,5	43,8	33,3	4,0	36,0	
autre	0,0	8,5	24,4	25,8	20,0	25,0	82,8	17,4	36,9	6,4	

Source : Insee, RP, A-2

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

TABLEAU 2c. – Les migrants interrégionaux (entre A-7 et A-2). Population arrivée dans la commune en provenance d'une autre région (% de la population de A-2)

caractéristiques en A-2	Communes									
	1	2	3	4	6	7	9	10	11	12
ensemble	18,2	16,8	20,4	10,7	9,1	2,9	9,7	9,2	9,0	10,6
Age en A-2										
moins de 40 ans	21,2	27,3	34,4	11,5	21,1	6,3	13,6	20,9	12,5	14,7
plus de 40ans	15,0	6,4	12,3	8,4	2,8	0,0	6,8	0,0	8,1	7,0
Sexe										
hommes	10,3	17,2	24,1	7,8	4,2	2,9	0,0	8,7	3,0	16,3
femmes	26,5	16,4	18,3	19,6	12,9	3,0	15,1	10,9	10,5	6,4
Statut occup										
logt en A-2										
propriétaire	40,2	14,7	24,6	9,2	9,1	0,0	12,3	9,0	6,4	12,3
locataire	16,7	24,1	16,1	12,1		5,4	0,0	9,9	11,7	8,6
Activité										
actif occupé	14,5	21,4	23,4	10,8	12,5	3,4	17,9	9,8	7,0	12,1
chômeur		66,7	0,0	0,0	0,0	0,0	0,0	0,0	11,1	27,5
retraité	8,0	0,0	17,2	13,6	6,3	0,0	0,0	0,0	6,9	1,9
étudiant	38,5	11,8	8,9	11,5	14,3	6,3	0,0	33,3	15,8	18,0
autre	19,2	8,5	35,9	6,5	0,0	0,0	0,0	0,0	16,5	3,5

Source Insee, RP, A 2

$$POP(t_2) - POP(t_1) = (N - D) + (I_{int} - E_{int}) + (I_{ext} - E_{ext}) \quad [1]$$

où : $POP(t_1)$ = population totale de la zone au 1^{er} recensement

$POP(t_2)$ = population totale de la zone au 2^e recensement

$(N - D)$ = solde naturel intercensitaire

$(I_{int} - E_{int})$ = solde migratoire intercensitaire interne

$(I_{ext} - E_{ext})$ = solde migratoire intercensitaire avec l'extérieur

Avec le RRP, on connaîtra, pour chaque zone du territoire, pour une période t_1 , t_2 telle que $t_2 = t_1 + 5$ (donc une période de 5 ans) :

– les populations en t_1 et t_2 (observations ou estimations)

– le solde naturel entre t_1 et t_2 (état civil),

– les entrées et les sorties en provenance et vers les autres zones de France métropolitaine (estimations réalisées à partir de la question sur le lieu de résidence 5 ans plus tôt),

– les entrées en provenance de l'étranger.

Les sorties vers l'étranger devraient théoriquement pouvoir être estimées à l'aide de la formule suivante :

$$E'_{ext} = I_{int} - E_{int} + I_{ext} - POP(t_2) + POP(t_1) + (N - D) \quad [2]$$

Pour la période 1982-1990, les estimations, pour chaque commune, du nombre de migrants vers l'extérieur conduisaient à un chiffre négatif dans 40 % des

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

TABLEAU 3. – Migrations entre les deux régions (entre A-7 et A-2)

	entrants				sortants				solde migratoire				pop totale	
	région 1		région2		région 1		région2		région 1		région2		région 1	
	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)	effectif	taux (%)
ensemble	71 5	11,6	173,7	28,2	173,7	28,2	71,5	11,6	-102,2	-16,6	102 2	16,6	615,0	615 0
Age en A-2														
moins de 40 ans	43 9	17,2	66,6	26,1	66,6	26,1	43,9	17,2	-22,8	-8,9	22,8	8,9	255 0	255,0
plus de 40 ans	27,6	7,7	107 0	29,7	107,0	29,7	27 6	7,7	-79,4	-22,1	79,4	22,1	360,0	360,0
Sexe														
hommes	26,9	9,1	87,3	29,6	87,3	29,6	26,9	9,1	-60,4	-20,5	60,4	20,5	285 0	285 0
femmes	44 6	13,9	86,4	27,0	86,4	27,0	44,6	13,9	-41,7	-13,0	41,7	13,0	320,0	320,0
Statut occup														
logé en A-2	38 7	11,0	101,7	29,1	101 7	29,1	38,7	11,0	-63,1	-18,0	63,1	18,0	350 2	350,2
propriétaire	32,8	12,4	71,9	27,2	71,9	27,2	32,8	12,4	-39 1	-14,8	39,1	14,8	264,8	264,8
locataire														
Activité en A-2														
actif occupé	35 4	12,8	81 9	29,8	81,9	29,8	35 4	12,8	-46,5	-16,9	46,5	16,9	275 2	275 2
chômeur	4 3	15,2	9,7	34,4	9,7	34,4	4,3	15,2	-5,4	-19,3	5 4	19,3	28 1	28 1
retraité	8 7	6,3	36,9	26,6	36,9	26,6	8 7	6,3	-28,1	-20,3	28,1	20,3	138 7	138,7
étudiant	15 7	14,3	29,0	26,5	29,0	26,5	15 7	14,3	-13,3	-12,2	13,3	12,2	109,4	109,4
autre	7 5	11,8	16,2	25,5	16,2	25,5	7,5	11,8	-8 7	-13,8	8,7	13,8	63 6	63,6

Source Insee, RP, A-2

PREMIÈRE APPROCHE DE L'ÉTUDE DES MIGRATIONS

communes. Au niveau des départements, on obtenait un chiffre négatif dans 16 départements.

L'utilisation des données censitaires pour estimer l'émigration extérieure et donc le solde migratoire avec l'extérieur de la métropole, au niveau local, semble peu fiable. L'estimation des sorties vers l'extérieur de la métropole cumule en effet des erreurs commises sur les chiffres des entrées internes, des sorties internes, des entrées externes (réponses à une question et donc soumises à un risque d'erreur), ainsi que celles affectant les dénombrements de population lors des recensements. Le problème sera identique avec le RRP.

Conclusion

La variable « *Lieu de résidence 5 ans auparavant* », sera, dans le RRP, une variable « comme les autres » : on applique à chaque individu un facteur d'ajustement obtenu par l'exploitation de sources administratives. Il en est de même pour les autres données bilocalisées (pour les navettes domicile- travail, en particulier, la variable « *lieu de travail* » sera également une variable « comme les autres »).

La réalisation et la diffusion de tableaux tels que ceux actuellement publiés par l'Insee, ou de tabulations sur mesure, ne devraient donc poser aucun problème.

Par ailleurs, le RRP apporte au moins deux avantages importants pour l'analyse de la mobilité et des migrations résidentielles :

- La possibilité de disposer d'évolutions annuelles
- Une question portant sur un intervalle de temps plus réduit (5 ans) et fixe.

RECENSEMENT RÉNOVÉ DE LA POPULATION COMPTE RENDU DES ÉCHANGES

Jean-Marie GROSBAS, Dominique ALLAIN,
Jean-Michel DURR et Alain GODINOT

1. Acceptabilité de la rénovation

Question : ne peut-on avoir des réactions d'élus locaux similaires à celles qui ont partagé les États-Unis entre Démocrates et Républicains il y a deux ans lorsqu'il était question d'utiliser les résultats d'un sondage pour la répartition des sièges au Congrès ?

Réponse : nous ne notons pour le moment aucune opposition de principe voire d'objections de la part des élus locaux alors qu'environ 2 000 d'entre eux ont été bien informés des dispositions envisagées. C'est une question qui n'est pas non plus apparue dans le débat à l'Assemblée Nationale. Nous faisons l'hypothèse qu'il en sera ainsi pour la suite.

Question : un participant considère que le risque de rejet de la rénovation existe car la concertation porte plus sur le détail technique que sur l'opportunité de la rénovation et les tests apparaissent un peu illusoire puisque réalisés après la prise de décision. Le doute sur la rénovation peut alors provenir de différents facteurs : des taux de réponse peu élevés pour les tests, des données annoncées plus fraîches qui se font attendre pour les premières prévues fin 2008 et enfin des résultats de qualité non avérée, difficiles à utiliser voire impossible à utiliser pour les analyses infracommunales. Au total, la réforme lui apparaît imposée alors qu'elle n'était pas nécessaire à ses yeux.

Réponse : c'est le législateur qui décidera, c'est le fondement même de la démocratie. La méthode proposée conduit effectivement à de premiers résultats légaux et à l'infracommunal, authentifiés pour les premiers et diffusés pour les seconds, fin 2008. Après cette montée en charge, les résultats seront produits tous les ans. Les tests débutés début 2000 se poursuivront en parallèle de la construction du dispositif juridique du RRP sachant que la loi est une loi de principe et que c'est dans le décret que sera précisé l'ensemble des modalités d'application ; d'où l'importance de réaliser des tests avec les communes. Il y a sur ces questions un intérêt partagé des communes et de l'Insee à ce que les choses se passent bien. Les taux de réponse aux tests sont similaires à ceux observés pour les tests du recensement de 1999. Ces taux

RECENSEMENT RÉNOVÉ DE LA POPULATION

sont issus d'une procédure qui, contrairement aux opérations en grandeur réelle, est basée sur le volontariat, sans relance, et avec une campagne de communication sommaire nettement inférieure. Quant aux données produites et à leur utilisation, la concertation se poursuivra sous l'égide du CNIS à la lumière des enseignements de la diffusion du RP99.

Question : le recensement traditionnel donnait des résultats basés sur un décompte exhaustif. On savait que cette notion d'exhaustivité était trompeuse. Maintenant, affichera-t-on des données avec des intervalles de confiance, avec le risque de contestation éventuel ?

Réponse : comme avant, les chiffres de population légale seront authentifiés par décret. Ces chiffres de dénombrement, publiés sans intervalle de confiance, s'appliqueront aux textes qui y font référence. On peut noter qu'avec des résultats annuels, les évolutions de population seront mieux prises en compte. D'ailleurs, les élus rencontrés nous ont semblé dans l'ensemble plutôt favorables à l'émergence d'un dispositif permettant de produire des résultats annuels proches de la « population réelle ». Pour les données statistiques, les intervalles de confiance seront disponibles.

Question : qu'est-ce qui demeure exhaustif dans le recensement rénové ? Le répertoire d'immeubles localisés (RIL), un répertoire de logements ?

Réponse : dans les communes de moins de 10 000 habitants, on procède à un recensement exhaustif tous les cinq ans. Dans les communes de 10 000 habitants ou plus, le RIL sera maintenu à jour en continu avec diverses procédures internes à l'Insee et grâce aux échanges avec les communes. Le RIL, qui est un répertoire d'adresses et non de logements, sera exhaustif.

2. Aspects juridiques du RRP

Question : des précisions sont demandées quant aux raisons qui motivent une rédaction très précise pour la transmission de données issues de l'assurance maladie dans le second alinéa du VII de l'article 59 ; la rédaction du reste de l'article apparaissant moins détaillée.

Réponse : la loi n° 51-711 du 7 juin 1951 modifiée sur l'obligation, la coordination et le secret en matière de statistiques régit l'ensemble des enquêtes statistiques des services publics ; elle s'applique au RRP. Pour expliquer la rédaction du second alinéa du VII, il convient de faire une mention explicite de l'article 7 bis de cette loi. Cet article permet aux personnes morales de droit public et aux personnes morales de droit privé gérant un service public de céder à l'INSEE, à des fins exclusives d'établissement de statistiques, des informations relatives aux personnes physiques et aux personnes morales recueillies dans le cadre de leur mission. Dans le cas des organismes gestionnaires des prestations de l'assurance maladie, le second alinéa évoqué rend cette transmission obligatoire.

Question : toujours sur cet article, comment doit-on interpréter que, pour les informations non nominatives extraites des fichiers sociaux et transmises à

l'Insee, la loi stipule qu'il appartient à l'Institut de les agréer, cinq ans après leur réception, à un niveau géographique de nature à éviter toute identification des personnes ?

Réponse : la CNIL s'est montrée soucieuse d'encadrer les dispositions liées à cet échange. Tout en prenant acte de l'intérêt que pouvait représenter pour l'Insee la possibilité d'utiliser des informations individuelles non nominatives, utilisables le cas échéant à l'adresse, elle a souhaité que l'Insee les agrège à l'issue d'un cycle de collecte, soit cinq ans après leur réception, à un niveau supprimant tout risque d'identification des personnes concernées.

Question : vous comptez établir les chiffres de population légale tous les ans même sans collecte ?

Réponse : le gouvernement, suivant en cela la recommandation du Conseil d'État dans son avis du 2 juillet 1998, a inscrit dans la loi le principe d'une authentification annuelle des chiffres de population légale des communes. Cela est une des garanties du respect du principe d'égalité entre les communes. Il appartiendra à l'Insee de déterminer chaque année les chiffres qui seront authentifiés.

3. Les questions méthodologiques

3.1. Le mode de sondage

Question : plusieurs intervenants se sont interrogés quant à l'effet du plan de sondage sur la variance, qu'il s'agisse du passage de l'immeuble à l'adresse comme unité pour la base de sondage, ou qu'il s'agisse du passage du tirage de logements au tirage d'adresses. Et cette question apparaissait d'autant plus importante que l'on ne gagne que deux à trois points de réponse en passant d'un mode liste à un mode adresse.

Réponse : le passage d'un sondage aléatoire à un sondage par grappe fait perdre en précision car on augmente la variance de l'estimateur. Entre un sondage par grappe et un sondage aléatoire simple, l'effet sur le coefficient de variation mesuré à l'Insee à partir de l'échantillon maître va de quatre à dix points. Mais cet effet n'est pas directement applicable au RRP puisque le mode de sondage introduit un nouveau facteur qui est l'équilibrage a priori des groupes de rotation dans lesquels on tire les échantillons. Au demeurant, pour juger de l'impact de la méthode sur la précision des estimateurs, la mesure du biais est tout aussi importante que la mesure de la variance. Or notre choix de passer d'un tirage de logements à un tirage d'adresses est fondé sur la réduction du biais de couverture avec le mode adresse. Le test a bien montré qu'il était particulièrement difficile d'être assuré en mode liste (tirage de logements) que la collecte se déroulerait bien dans le logement tiré et non dans un logement à côté. Recenser tous les logements d'une adresse rend le dispositif plus robuste.

3.2. La non réponse dans les tests, ses facteurs explicatifs et les perspectives

Question : plusieurs intervenants reprenant les facteurs influençant la réponse ont souhaité obtenir des précisions sur « l'effet régional » dont le poids est important dans l'analyse.

Réponse : le plan d'expérience a été construit pour mesurer le différentiel des taux de réponse entre le mode de sondage liste de logements et le mode de sondage adresse et, pour ce facteur, l'effet est faible : on gagne 2 à 3 points avec le mode adresse. Lors de l'exploitation, l'Insee a tiré parti de ce test et a introduit dans l'analyse de la non réponse d'autres facteurs explicatifs, à savoir le type de bâti et l'effet régional. Ce dernier combine de nombreux facteurs qui vont de l'organisation du travail et de son suivi à l'effet enquêteur. Pour l'effet régional, l'analyse s'est concentrée sur l'effet enquêteur en prenant plusieurs descripteurs (sexe, âge, niveau de diplôme, ayant travaillé avant, etc.). Mais il n'a pas été possible de réaliser dans chaque région tous les croisements possibles.

Question : il existe une probabilité importante que la non réponse soit concentrée sur les ménages d'une personne; les caractéristiques des non répondants sont donc à étudier tout particulièrement pour alimenter le processus d'estimation.

Réponse : on pourra poursuivre l'analyse prévue pour les tests en intégrant, dans la mesure du possible, des variables socio-démographiques.

3.3. Les données produites par le RRP

Question : certains intervenants demandent des précisions sur le décalage temporel envisagé dans la méthode d'estimation des données détaillées présentée en séance.

Réponse : sur le cycle de collecte de cinq ans, il s'agit tout d'abord d'utiliser les collectes de chaque année et de les ramener à une même date. Pour une production de données en A, année de fin de cycle, c'est l'année médiane du cycle, en A-2, qui est utilisée pour synchroniser les collectes, ce qui a l'avantage de limiter à la durée la plus courte possible les extrapolations ou rétroprojections nécessaires. Pour guider ces extrapolations et rétroprojections, les évolutions observées de données extraites des fichiers administratifs disponibles seront utilisées.

Question : qu'en est-il de la méthode si les données administratives nécessaires ne sont pas disponibles, par exemple pour l'année A alors que l'on voudrait rétroprojectionner les données observées en A sur A-2 ?

Réponse : dans ce cas, faute de taux d'évolution de données administratives, il est possible de recourir à des techniques d'interpolation linéaire.

Question : comment prendrez-vous en compte l'amélioration des fichiers administratifs dans votre processus d'estimation ? Il y aura bien des phases

RECENSEMENT RÉNOVÉ DE LA POPULATION

d'apurement de doubles comptes avec baisse consécutive des effectifs, comment seront-elles gérées ?

Réponse : les fichiers sociaux ne sont pas utilisés en niveau mais peuvent contribuer à estimer les évolutions ; ce qui suppose une analyse préalable de leur qualité.

Question : les géographes veulent mettre en évidence des ruptures spatiales. Avec une perte de précision spatiale liée au sondage, des fichiers détail non disponibles et un processus d'estimation qui empile des données dans les villes, comment détecter les ruptures ?

Réponse : en ce qui concerne les communes qui feront l'objet d'enquêtes de recensement par sondage, l'inertie liée au modèle d'estimation est effectivement à étudier plus précisément. Mais les ruptures étaient encore moins appréhendables tout au long de l'espace de temps croissant qui séparait deux recensements traditionnels. Avec le recensement rénové, les ruptures seront plus rapidement perçues mais leurs effets seront lissés dans le temps.

Question : on peut considérer que la méthode d'estimation présentée s'adapte bien aux communes de moins de 10 000 habitants. Mais, pour les autres communes, où il y a de la collecte tous les ans, est-ce la méthode optimale ?

Réponse : il faudra effectivement optimiser les différents estimateurs. La méthode présentée peut se transposer aux communes de 10 000 habitants ou plus. Pour ces communes, des travaux portant sur la mise au point de méthodes élaborées à partir de la combinaison de cinq échantillons successifs de collecte sont envisagés.

Question : y a-t-il nécessité de caler les estimations détaillées issues du cycle de collecte sur cinq ans aux estimations issues du seul sondage annuel ? Ne peut-on envisager d'autres sources de calage ou autre chose ?

Réponse : les sondages de l'année donnent des estimations fiables aux niveaux national et régional. À ce titre, elles constituent une référence pour les estimations détaillées produites deux ans plus tard. La question de mettre en cohérence les deux séries se pose, ne serait-ce que dans un souci de simplification pour l'utilisateur. Il faut cependant étudier soigneusement les conséquences d'une telle opération au niveau communal. C'est donc au vu de l'ensemble des objectifs d'un recensement que cette question peut être tranchée.

Question : Vous aurez des pondérations différentes entre les logements et les individus, comment comptez-vous assurer la cohérence de ces deux séries ?

Réponse : C'est une question qui est classique dans les enquêtes et on appliquera les méthodes généralement utilisées pour la traiter.

Question : Pour les migrations, comment construire la matrice inversée, comment repérer ceux qui sont partis ?

Réponse : Le fichier de référence d'une année sera constitué des logements et des individus recensés cette année-là et les quatre années l'entourant,

RECENSEMENT RÉNOVÉ DE LA POPULATION

pondérés pour tenir compte des évolutions. Ce fichier représente donc ce qu'aurait donné un fichier issu d'observations de la seule année concernée. Ainsi, chaque information obtenue à partir de ce fichier estimera l'information correspondante de l'année. Il en va des flux migratoires comme des autres caractéristiques socio-démographiques, comme l'a montré B. Baccaïni dans son exposé.

Question : comment construire les tableaux bilocalisés, comment est gérée la rupture du mode de collecte entre les communes ?

Réponse : Comme indiqué précédemment, pour l'utilisateur la rupture du mode de collecte entre les communes est « transparente ». Une année donnée, il disposera d'un fichier représentant ce qu'aurait donné un fichier issu d'observations de la seule année concernée. Prenons, par exemple, le cas des données domicile-travail. Chaque année on a la « destination » des personnes interrogées dans une commune émettrice donnée. Il est donc possible d'avoir un estimateur pour les communes réceptrices : on est dans la problématique de l'estimation sur domaine. En régime de croisière, les données recueillies en cinq ans permettent d'avoir des estimateurs consolidés.

Question : Ce que l'utilisateur souhaite connaître, c'est l'influence du taux de sondage sur les événements qu'il veut mesurer et l'inertie de la mesure de phénomènes ponctuels alors que le modèle d'estimation est basé sur un cycle de cinq ans.

Réponse : on ne peut répondre à cette question dans l'absolu car tout dépend du phénomène à mesurer et de sa fréquence d'observation. Les effets seront plus importants sur des observations rares.

Question : des précisions sont demandées suite à la présentation de l'étude des migrations avec le RRP. Le mécanisme d'estimation présenté consiste-t-il à transposer d'un an les observations collectées modulo un facteur d'ajustement pour passer de A-8/A-3 à A-7/A-2 ?

Réponse : l'exemple présenté pour les migrations permet d'illustrer un mécanisme de production des estimations détaillées. Les données de l'exemple ne sont pas réelles, mais ont été construites pour voir la façon dont le modèle d'estimation réagit. On a d'abord créé un fichier pour des populations communales calquées sur des structures types du RP 1999, y compris sur les mouvements migratoires. On a appliqué à ces données le modèle d'estimation avec des données administratives fictives ayant des évolutions très irrégulières. En sortie de modèle, les estimations à la commune ont les bonnes structures, mouvements migratoires compris. Pour répondre précisément à la question posée, en A-2, le fichier de sortie est un fichier détail dans lequel chaque ligne est une unité avec ses variables collectées au moment où elle a été enquêtée avec une pondération qui :

- tient compte du taux de sondage ;
- est modulée, pour les unités non enquêtées en A-2, par les taux d'évolution de données administratives.

Ainsi, pour les unités enquêtées en A-3, le taux d'évolution sera de A-2/A-3, y compris pour la réponse à la question sur la résidence antérieure à cinq ans (i.e. en A-8). Ce sont les pondérations qui changent en fait et non les données.

Question : Un effort sur la pédagogie pour expliquer la méthode et les simulations présentées est demandé.

Réponse : c'est une recommandation qui nous a été faite par le comité scientifique en 2000 et que nous entendons. La méthode peut paraître assez compliquée pour l'heure et nous avons effectivement à entreprendre des efforts pour la rendre totalement compréhensible.

4. Organisation et mise en œuvre : les moyens et les procédures

Question : l'Insee présente un recensement rénové à coût constant ; la Nation ne peut-elle se situer dans un contexte où le budget augmenterait ?

Réponse : lors de la négociation budgétaire, l'Insee a obtenu un accord sur une enveloppe qui permet d'envisager la rénovation à un coût similaire à celui des recensements généraux. Il n'est pas envisageable de demander une rénovation à coût croissant.

Question : Quand et comment la mise à jour du RIL est-elle prévue ? La concertation et le partenariat avec les communes sont indispensables car la loi va fixer le principe des échanges.

Réponse : le calendrier et la procédure pour parvenir à un RIL 1999 sont fixés. Au printemps 2002, date de constitution interne du RIL 1999, les échanges entre les directions régionales de l'Insee et les communes se déploieront. Les systèmes cartographiques étant divers, il faudra prévoir différentes possibilités d'échanges dans le cadre d'une convention type.

Question : Pourquoi avoir testé un prérecensement alors qu'on le faisait déjà depuis 1990 et ne peut-on augmenter le taux de sondage avec le passage de la liste à l'adresse ?

Réponse : ce qui existait en 1990 et 1999 étaient des tournées de reconnaissance et non des prérecensements ; les objectifs et conditions de réalisation ne sont pas comparables. Quant à l'augmentation du taux de sondage, l'Insee a pour le RRP un budget qui s'impose à lui. A l'intérieur de cette enveloppe, les crédits seront distribués et optimisés pour la rénovation.

Question : l'analyse des erreurs de déclaration dans les catégories de logement n'a pas été évoquée, est-ce traité ?

Réponse : les premières analyses réalisées à l'occasion des tests effectués au printemps 2001 donnent un taux d'erreur de 1 %, similaire aux taux des tests réalisés pour les recensements généraux (de l'ordre de 1,3 % d'après l'enquête post-censitaire de 1990).

Question : l'analyse des facteurs influençant la réponse montre toute l'importance de l'agent recenseur dans le dispositif, mais aussi de son encadrement. Cette professionnalisation a un coût.

Réponse : l'agent recenseur est important quant à la qualité de la collecte, c'est évident. L'apprentissage du travail qui peut conduire à une professionnalisation accrue passe tout d'abord par la formation. L'Insee réfléchit au contenu et à l'organisation de la formation des agents recenseurs et associera les communes à cette réflexion. Par ailleurs, l'Insee prévoit de dispenser un module de formation aux responsables communaux en charge du suivi du travail de l'agent recenseur. Il est rappelé que, pour les tests, il avait été expressément demandé que les enquêteurs n'aient aucune expérience passée en tant qu'agent recenseur.

Question : l'Insee indique que les taux de réponse augmenteront dès lors qu'il y aura une bonne campagne de communication, mais l'efficacité de cette campagne reste à prouver.

Réponse : la communication est essentielle pour un recensement et l'on pressent bien que la communication de proximité devra être développée même si une communication nationale reste nécessaire, surtout au moment de l'initialisation. Remarque : un participant considère que la communication ne pourra plus se dérouler comme avant mais que si l'on avance de pair sur l'angle communication et l'angle professionnalisation, il est possible d'en tirer parti.

5. Autres thèmes : contenu de l'information, transfert de méthode à d'autres pays

Question : le concept de résidence principale des recensements devient de plus en plus éloigné des modes de vie ; comment traitez-vous le développement de la multi résidence ?

Réponse : nous avons testé une question sur la multi résidence dans les tests réalisés en 2001. Les tests montrent qu'on n'a pas trouvé encore de formulation permettant d'atteindre les objectifs statistiques visés. C'est pourquoi cette question ne peut pas être introduite dans les premières campagnes de collecte.

Question : la rénovation permet dans l'ensemble d'alléger la charge à supporter pour les recensements de population. Ne pourrait-elle être appliquée dans des pays qui ne peuvent supporter cette charge comme les pays en voie de développement ?

Réponse : cette question est à traiter dans les structures d'échanges entre pays.

SYNTHÈSE DES DÉBATS

Henri LERIDON *

En rappelant la richesse des définitions apportées au cours des temps au terme de synthèse, et notamment celle de Kant qui veut que la synthèse ajoute à une représentation un point de vue, je crois que la variance de ceux qui furent entendus à ce séminaire apporte quelques degrés de liberté pour cet exercice.

1. Les évolutions depuis le séminaire de l'automne 2000

Trois éléments et décisions notables depuis le dernier séminaire ont été rappelés :

- 1) Les enquêtes sont réalisées en une seule vague de collecte, ce qui présente un intérêt certain pour les questions de professionnalisation, les aspects techniques étant à apprécier ;
- 2) Le passage de l'immeuble à l'adresse et, ce faisant, la décision de limiter le risque de biais de couverture ;
- 3) Les tests sur la non-réponse et l'effet du mode de sondage dont l'analyse a été présentée aujourd'hui.

Ces expérimentations et évolutions sont à mettre à l'actif du recensement rénové mais il n'en demeure pas moins des questions en suspens.

2. Les questions que suscite la méthode

Cinq questions ont été plus particulièrement développées :

- 1) Les questions liées à la qualité des fichiers administratifs et à l'effet de leur utilisation : s'il n'y a pas de risque important sur l'évolution et le changement de nature des fichiers de la taxe d'habitation, il sera en revanche impossible de tester les fichiers sociaux, leur évolution sera vraisemblablement source de gêne et il est peu probable qu'ils seront mis à jour rapidement, la logique de gestion n'apparaissant pas y concourir. Par ailleurs, l'effet de leur imprécision sur les estimations détaillées est à apprécier.
- 2) Les variables sexe et âge sont majeures dans la méthode, que ce soit pour équilibrer a priori les groupes de rotation ou lors de l'utilisation de l'évolution de données administratives. Comment réagissent les autres variables ?
- 3) La question de la collecte et celle de son contrôle sont à étudier aussi tout particulièrement.

* INED, 133, Boulevard Davout, 75980 Paris Cedex 20, E-mail leridon@ined.fr

SYNTHÈSE DES DÉBATS

4) Les modes de vie évoluent, la résidence multiple ou la double résidence se développent, la résidence principale peut devenir ambivalente. Comment résoudre cette question avec le recensement rénové ?

5) Enfin, bien que la question ait été peu évoquée dans le séminaire, la base du recensement de 1999 s'éloigne, alors que cette base aurait pu être utilisée comme calage ; quelles en sont les conséquences pour l'initialisation du recensement rénové ?

En conclusion, on peut souhaiter trouver un équilibre entre la nécessaire rénovation et les risques de discontinuité liés au changement de formule. Partant de l'idée que, jusqu'au démarrage effectif du système tournant, le RP reposait implicitement sur le principe de « tickets financiers annuels » que l'on accumulait jusqu'à la réalisation d'une opération de recensement, je proposerais que l'on utilise le capital engrangé depuis 1999 en 2004, en réalisant cette année-là une opération intégrant les grands principes du recensement rénové, sauf l'aspect « tournant ». L'opération pourrait être assimilée à une répétition générale, permettrait de faire patienter pendant l'attente des premières populations légales issues du nouveau système, et faciliterait l'acceptation de ce système par les acteurs politiques.

BIBLIOGRAPHIE

- BACCAÏNI B. (1999), « Analyse des migrations internes et estimation du solde migratoire externe au niveau local à l'aide des données censitaires ». *Population*, n° 4-5, 801-815.
- BACCAÏNI B. (1999), « Recensement traditionnel et recensement rénové pour l'étude des migrations résidentielles ». INSEE, 101 p.
- BERTRAND P. (2000), *Estimations annuelles dans la rénovation du recensement de la population*, note de travail interne, Département de la démographie, INSEE.
- BORCHSENIUS L. (2000), « From a Conventional to a Register-based Census of Population », Les Recensements après 2001, Séminaire Eurostat,-INSEE, Paris.
- DEVILLE J.C., TILLÉ Y. (1999), *Balanced Sampling by Means of the Cube Method*, CREST-ENSAI, document interne, soumis pour publication.
- DEVILLE J.C., TILLÉ Y. (2000), « Échantillonnage équilibré par la méthode du cube et estimation de variance », *Journées de Méthodologie*, décembre 2000, INSEE, Paris.
- HORVITZ D.G. (1986), « Statement to the Subcommittee on Census and Population », Committee on Post Office and Civil Service, House of Representatives, Research Triangle Park, North Carolina.
- JACOD M., DEVILLE J.-C. (1996), « Replacing the Traditional French Census by a Large Scale Continuous Population Survey », *Annual Research Conference Proceedings*, USBC, Washington.
- KISH L. (1981), « Population Counts from Cumulated Samples », Congressional Research Service, *Using Cumulated Rolling Samples to Integrate Census and Survey Operations of the Census Bureau*, Prepared for the Subcommittee on Census and Population, Committee on Post Office and Civil Service, House of Representatives, Washington.
- KISH L. (1990), « Recensement par étapes et échantillons avec renouvellement complet », *Techniques d'enquêtes*, Vol 16, N° 1, pp. 67-86, Statistique Canada, Ottawa, juin 1990.
- KAUFFMANN B. (2000), *Estimation de la précision due au modèle de synthèse*, note de travail interne, Département de la démographie, INSEE.
- LAIHONEN A. (2000), « 2001 Round Population Censuses in Europe », *Les Recensements après 2001*, Séminaire Eurostat-INSEE, Paris.
- NATHAN G. (2001), « Models for combining longitudinal data from administrative sources and panel surveys », Présentation invitée, ISI, Séoul, Août 2001.

BIBLIOGRAPHIE

- ONU (1990), *Principes et recommandations complémentaires concernant les recensements de la population et de l'habitat*, Études statistiques, ST/ESA/STA/sérieM/67, New York.
- SÄRNDAL C.E. (1990), « Méthodes pour estimer la précision des estimations d'enquête lorsqu'il y a eu imputation », *Recueil du Symposium 90 de Statistique Canada : Mesure et amélioration de la qualité des données*, Ottawa, octobre 1990, pages 369-380.
- (1994) « Radical Alternatives », *Modernizing the U.S. Census*, B. Edmonston et C. Schultze, éditeurs ; Panel on Census Requirements in the Year 2000 and Beyond, National Research Council, National Academy Press, pp. 59-74.