Recueil du Symposium 2001 de Statistique Canada
La qualité des données d'un organisme statistique : une perspective méthodologique

LA RÉNOVATION DU RECENSEMENT FRANÇAIS

Jean-Michel Durr[1], Jean Dumais[2]

RÉSUMÉ

Il devient de plus en plus difficile de mener des recensements de façon traditionnelle. Quand on a la possibilité d'interconnecter des fichiers administratifs, on ouvre une alternative intéressante à la pratique de recensements périodiques (Laihonen, 2000 ; Borchsenius, 2000). On retrouve ce type de proposition dans un article récent de Nathan (2001). La rénovation développée à l'INSEE repose sur le concept de « recensement continu » dont l'idée remonte à Kish (1981, 1990) et Horvitz (1986). Une première approche envisageable en France peut être trouvée dans Deville et Jacod (1996). Le présent article fait le point des développements méthodologiques depuis que l'INSEE a mis en route son Programme de rénovation du recensement de la population.

MOTS CLES : échantillonnage équilibré, recensement, recensement continu, calage

1. INTRODUCTION

1.1. Les raisons de la rénovation

La France conduit depuis de nombreuses années des recensements afin de déterminer la population légale de chacune de ses circonscriptions administratives et décrire les caractéristiques socio-démographiques de ses territoires à tous les niveaux géographiques, des quartiers des communes au pays dans son ensemble. Ainsi, le recensement de 1999 s'est déroulé selon le schéma habituel. Toutefois, certains éléments nous ont conduits à revoir ce dispositif. Tout d'abord, l'intervalle intercensitaire a tendance à s'allonger pour des raisons budgétaires. Ainsi est-on passé de recensements quinquennaux avant la guerre, à des écarts entre les recensements de 7, puis 8 et enfin 9 ans entre le recensement de 1990 et celui de 1999. De plus, le public ne comprend pas toujours la nécessité d'une opération aussi lourde dans un contexte de fichiers administratifs toujours plus nombreux, même s'il redoute par ailleurs leur utilisation croisée. Enfin, le mouvement de décentralisation que connaît la France depuis plus de vingt ans a généré de nombreux besoins de données statistiques afin d'éclairer les politiques locales. Le recensement, source d'information locale par excellence, doit donc s'adapter et fournir des données plus fraîches et toujours finement localisées.

C'est pourquoi un programme de rénovation du recensement de la population a été engagé à l'Insee dès la fin des années 90. La France ne disposant pas de registre de population et le contexte national ne s'y prêtant pas, il a donc été décidé d'envisager une voie intermédiaire combinant la réalisation d'enquêtes annuelles par sondage et l'utilisation de fichiers administratifs non nominatifs que l'Insee est habilité à utiliser à des fins exclusivement statistiques. Pour les communes dont la population est inférieure à un seuil, fixé pour l'instant à 10 000 habitants, les enquêtes seront exhaustives et auront lieu chaque année par roulement au cours d'une période de cinq ans. Pour les autres communes, une enquête par sondage sera effectuée chaque année, la totalité du territoire de ces communes étant prise en compte au terme de la même période de cinq ans. Pour mener à bien cette rénovation, un cadre juridique nouveau s'est avéré nécessaire. Le Conseil d'Etat, consulté sur le projet, a préconisé, dans son avis du 2 juillet 1998, que le gouvernement soumette au Parlement un

[1] Jean-Michel Durr, Programme de rénovation du recensement de la population, INSEE, Direction générale, 18 boul. Adolphe Pinard, 75675 Paris CEDEX 14, France
[2] Jean Dumais, Programme de rénovation du recensement de la population, INSEE, Rhône Alpes, 165 Garibaldi, 69401 Lyon CEDEX 3, France

projet de loi. Outre la nécessité de donner une assise légale au recensement, il a considéré que le changement important des modalités d'élaboration des chiffres de population, alors que plus de 200 textes législatifs ou réglementaires s'y réfèrent, nécessitait de passer par la voie législative. Dans le cadre ainsi défini, le projet de loi vise essentiellement à définir les principes et à fixer les règles de base applicables à l'organisation du recensement.

L'opération est placée sous la responsabilité et le contrôle de l'Etat : l'Insee organise le cadre de la collecte (concepts, protocoles), réalise le tirage des échantillons, veille à la qualité des informations collectées, exploite les données et les diffuse. Les communes ou leurs groupements préparent et réalisent les enquêtes de recensement. En compensation, l'Etat leur verse une dotation financière.

1.2. Objectifs de qualité

Les **objectifs de qualité** suivants sont assignés au Programme :

1.2.1. Qualité des données :

Actualité : L'objectif est de pouvoir diffuser avant la fin de chaque année A la population légale de toutes les circonscriptions administratives au 1er janvier A-2, une description statistique de toutes les zones du territoire (communes et groupements de communes, quartiers des grandes villes, pays, etc.), relatifs au 1er janvier A-2 et une description statistique au 1er janvier de l'année A pour la France et ses grands territoires (régions…). Par comparaison avec un recensement général, le recensement rénové fournira des résultats analogues sur la population et les logements pour un gain en termes de fraîcheur moyenne des données de l'ordre de 3 à 4 ans.

Pertinence : Les données produites doivent être pertinentes pour une utilisation locale. En particulier, les données dont l'étude ne se justifie qu'à des niveaux géographiques très supérieurs à la commune seront écartées au profit de données plus utiles à l'action locale. Le choix des données à collecter est effectué dans le cadre du Conseil national de l'information statistique (CNIS), qui comprend des représentants des différentes catégories de producteurs et d'utilisateurs des données statistiques publiques. Un groupe de travail du CNIS a proposé des évolutions tout en préservant une nécessaire continuité avec les recensements précédents et en limitant la charge de réponse.

Précision : Le recensement a pour objet de fournir des données significatives pour tous les niveaux géographiques du pays. Les données produites doivent avoir une précision suffisante, y compris aux niveaux infra-communaux, pour les répartitions les plus utiles à ces niveaux. Ceci concerne, en particulier, les structures par sexe et âge, par type d'activité, catégorie socioprofessionnelle et par catégorie de logement. La précision des données devra pouvoir être estimée et indiquée aux utilisateurs.

Compréhensibilité : Les données produites devront être compréhensibles aisément et d'utilisation comparable à celle des données produites par un recensement général, afin de ne pas perturber les utilisateurs.

1.2.2. Qualité du processus :

Charge de réponse : Afin de limiter la charge de réponse pour la population, les informations collectées devront se limiter au strict nécessaire. En particulier, les informations disponibles aux mêmes échelons géographiques dans d'autres sources ne seront pas collectées, sauf si l'obtention de ces informations dans le cadre du recensement permet des croisements utiles avec les autres variables. Le questionnaire individuel se limitera à un recto-verso, à l'instar des recensements précédents.

Questionnaire : la méthode de collecte étant basée sur le dépôt-retrait, les questionnaires doivent être accessibles à l'ensemble de la population. Afin de s'assurer de la compréhension des questions, des tests

qualitatifs, par la méthode des groupes de discussion, ont été menés. Par ailleurs, un test de collecte a été réalisé au premier semestre 2001 auprès de 4 000 logements.

Confidentialité : Les données mobilisées lors du recensement sont protégées par la loi. Les informations individuelles collectées ne doivent être accessibles qu'aux seules personnes autorisées. Ces informations sont destinées à l'Insee et ne peuvent être utilisées qu'à des fins exclusivement statistiques. Seules les informations indispensables à la préparation et à la réalisation des enquêtes de recensement sont partagées, pour ce qui les concerne, avec les communes ou leurs groupements.

Robustesse technique et organisationnelle : Compte tenu de l'ampleur des volumes traités et de l'enjeu attaché au recensement, le Programme doit s'appuyer sur des innovations techniques maîtrisées. En corollaire, le démarrage de l'opération devra privilégier la robustesse du dispositif. L'introduction d'innovations techniques ou fonctionnelles pourra se faire tout au long de la vie du recensement dans le cadre de la maintenance évolutive ou de projets spécifiques. Ces projets pourront ainsi s'appuyer sur les opérations annuelles pour tester leurs apports avant généralisation. Cependant, la notion de cycle quinquennal sera privilégiée pour l'introduction de modifications sensibles, telles que des évolutions des questionnaires. L'organisation doit s'appuyer sur un partenariat équilibré entre l'Insee et les communes. L'organisation envisagée devra être réalisable par l'Insee compte tenu de ses moyens et de son programme de travail, moyennant des aménagements de l'organisation de la production de l'Institut. De même, l'organisation doit être supportable par les communes ou les établissements de coopération intercommunale. Le rythme annuel en grande commune et la possibilité pour les petites communes de déléguer la collecte sur leur territoire à une structure intercommunale sont de nature à favoriser la professionnalisation des acteurs de la collecte.

Pour l'Insee, l'intégration des opérations de recensement dans le programme de travail annuel des directions régionales, de même que le volume de l'opération, sept fois moins important que celui du recensement général, permettra un contrôle plus approfondi de l'opération. En effet, au lieu de 60 millions d'individus à recenser dans 36 700 communes avec 110 000 agents recenseurs une année donnée, il n'y aura chaque année que 18 000 agents recenseurs qui rendront visite à près de 9 millions d'habitants, et ce dans 8 000 communes environ.

Afin de définir précisément le cadre d'organisation entre l'Insee et les communes, un décret prévoira les responsabilités respectives, les moyens à mettre en place par les communes ainsi que les processus de validation des différentes étapes.

Maîtrise des coûts : la réalisation de la collecte sur un cycle de 5 ans permettra d'étaler les charges financières affectées à l'opération. En ce qui concerne les communes de plus de 10 000 habitants, la charge du recensement rénové sera plus faible que celle d'un recensement général de population en régime courant. En revanche, la charge devrait, pour les communes de moins de 10 000 habitants, être identique à celle d'un recensement général, mais tous les 5 ans alors que la périodicité d'un recensement général était de l'ordre de 8 ans. Le coût du recensement rénové en régime permanent devrait être inférieur à 30,5 millions € (euros 2000) par an. Cependant, les premières années de collecte pourront supporter un budget quelque peu supérieur à ce montant afin de permettre le rodage du processus.

2. STRATÉGIE D'ÉCHANTILLONNAGE

La commune constitue le point d'ancrage de la rénovation : les «petites et moyennes communes » (celles de moins de 10,000 habitants) seront sondées au taux (moyen) d'un cinquième par an et tous leurs logements seront visités ; toutes les «grandes communes » seront visitées chaque année, mais seulement une fraction de leurs logements sera enquêtée.

2.1. Les petites et moyennes communes

Considérons d'abord le domaine des «petites et moyennes communes ». Dans chaque région, cinq groupes de rotation de communes seront formés. Ces groupes de rotation seront créés à partir des renseignements du recensement de la population (RP99) par tirage d'échantillons équilibrés (Deville, Tillé, (1999, 2000)) sur la distribution âge-sexe des communes ; cette approche devrait permettre de minimiser les variations interannuelles dues au seul sondage.

Les figures 1 et 2 illustrent comment les 5 groupes de rotation sont équilibrés. Ces deux figures donnent les « diagrammes à moustaches » de deux variables mesurées sur les 2811 petites communes de Rhône-Alpes au Recensement général de la population de 1990. Pour chaque groupe de rotation, on voit et les quartiles et l'étendue de la distribution; il est intéressant de noter la superposition des diagrammes. La variable « Nombre de femmes âgées de 20 à 39 ans » a été utilisée pour la composition des groupes; le nombre de résidences principales, ni aucune des variables associées au ménage ou au logement, n'intervient pas dans l'établissement de l'équilibre.

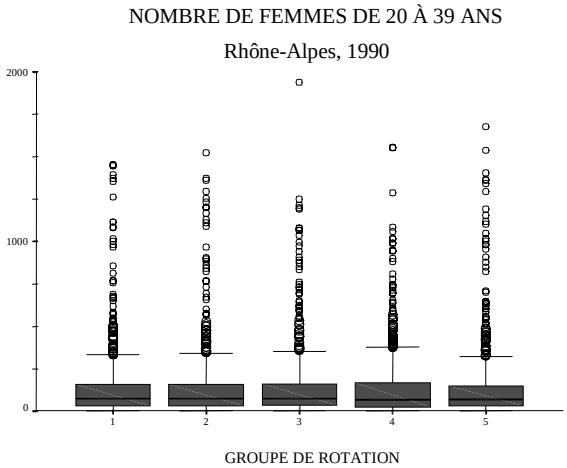

Figure 1

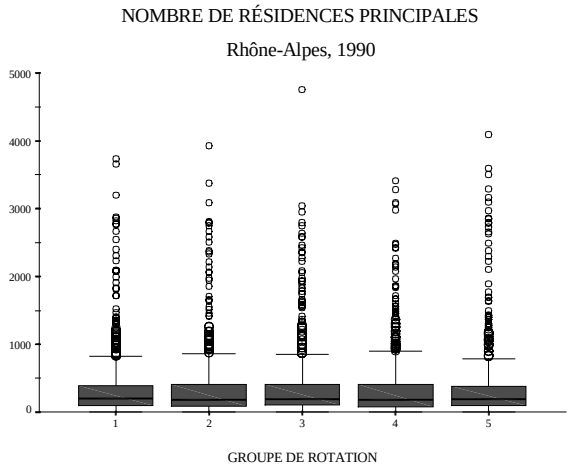

Figure 2

Chaque année, on fera le recensement (exhaustif) de la population et des logements de toutes les communes d'un des groupes de rotation. Ainsi, chaque «petite commune » sera recensée une fois tous les 5 ans, et toutes les «petites communes » à raison d'un cinquième par année.

2.2. Les grandes communes

Le sondage en «grande commune » utilisera le «répertoire d'immeubles localisés » (RIL). Ce répertoire est une liste d'édifices (résidentiels, institutionnels ou commerciaux) repérés individuellement de façon à créer une cartographie numérisée. Le RIL sera d'abord alimenté par les résultats du RP99 permettant ainsi de décrire statistiquement chaque immeuble résidentiel[3].

Le RIL sera mis à jour en continu à partir de permis de construire, de permis de démolir, de fichiers d'abonnés (eau, gaz, électricité, etc.), de renseignements fournis par les administrations locales et par l'observation directe sur le terrain. Ainsi, le RIL peut servir à la constitution d'une base de sondage «immeubles » en «grande commune ».

Dans chaque IRIS2000[4] de chaque «grande commune », on créera 5 groupes de rotation d'adresses sur le modèle du sondage en «petite commune ». Trois strates supplémentaires seront prévues dans chaque

[3] Au RP99, un «immeuble » est l'ensemble des logements desservis par la même cage d'escalier ; un même immeuble physique peut devenir plusieurs «immeubles du R.P. ».
[4] IRIS2000 = "*îlots regroupés selon des indicateurs statistiques*" zone homogène d'environ 2000 habitants

IRIS2000 : une pour les immeubles d'activité (usines, entrepôts,…), une seconde pour les logements collectifs (établissements, collectivités, communautés, internats,...), et une dernière pour les adresses neuves. On visitera chaque année un cinquième des immeubles d'activité pour s'assurer qu'ils sont toujours vides de logements (logement de gardien, ou espace converti à l'habitation) ; les logements éventuellement trouvés dans de tels immeubles seraient considérés autoreprésentatifs parce qu'exceptionnels. L'ensemble des logements collectifs sera couvert chaque année ; un cinquième d'entre eux seront visités alors que l'effectif des quatre autres cinquièmes sera éventuellement mis à jour par enquête téléphonique. Finalement, les immeubles d'habitation neufs seront recensés afin de pouvoir en faire le profil statistique et les insérer dans l'un des groupes de rotation.

Comme décrit plus haut, les groupes de rotation d'adresses seront visités à tour de rôle au cours d'une période de 5 ans. La collecte dans un groupe de rotation se fera selon l'une des deux options suivantes :

- pour les adresses d'habitation du groupe de rotation de l'année en cours, on dressera la liste des logements qu'ils contiennent dont on tirera un échantillon au taux de 40% ; les logements ainsi choisis seront invités à participer au recensement de l'année.
- un sous-échantillon d'une adresse sur deux sera tiré de la liste des adresses du groupe de rotation de l'année ; on a ainsi un taux de sondage de 50% des logements en moyenne.

En «grande commune », sur demande de la commune, on aura la possibilité d'augmenter l'échantillon de ménages jusqu'à couvrir 100% des logements du groupe de rotation.

En résumé, l'échantillon annuel comptera environ 8 millions de bulletins individuels, 6 millions des «petites communes » et 2 millions des «grandes communes ».

3. ESTIMATIONS GLOBALES ET DETAILLEES

En régime courant, trois séries d'estimations seront produites et diffusées chaque année : une série d'estimations des populations légales, une série d'estimations détaillées –dont on tirera les populations légales- et une série d'estimations globales servant au calage des précédentes.

3.1. Estimations globales

A ce jour, les plans de diffusion prévoient la publication, au 31 décembre *A*, des résultats nationaux et régionaux pour l'enquête tenue en début d'année *A* ; ces estimations forment les estimations globales pour l'année *A*. Il est aussi prévu de publier à la même date les résultats pour chacune des «petites communes » visitées durant la campagne de collecte de l'année *A*.

3.2. Estimations détaillées

Les fichiers administratifs fourniront des informations complémentaires à un niveau de détail assez fin. Il sera alors possible de mesurer la distorsion entre ce qui a été observé et ce qui est inscrit au fichier pour des objets similaires (immeubles, îlots,...). Cette distorsion sur des agrégats bien déterminés peut se traduire en facteur de correction à appliquer aux données administratives de sorte que la somme corrigée de celles-ci corresponde bien aux estimations censitaires.

A ce jour, il est prévu d'exploiter les fichiers administratifs à un niveau d'agrégation géographique (immeuble, îlot, district d'agent recenseur,...) qui renseigne sur les individus (âge, sexe d'après les fichiers de l'assurance maladie) ou leurs logements (fichiers de taxe d'habitation, ci-dessous *TH*).

Les résultats détaillés relatifs à l'année *A-2* seront mis à disposition au 31 décembre *A*[5]; ces résultats détaillés seront l'amalgame d'observations faites par sondage (en grande commune) ou recensement (en petite commune) et de données synthétiques.

[5] Il est prévu que l'acquisition et le traitement des fichiers administratifs prendront environ 2 ans.

Les données synthétiques seront obtenues à partir de la relation entre données observées et administratives sur un même point en un même instant. Par exemple, pour une commune *C* du Groupe II recensée en *A-3*, dont le recensement est établi à $R_{C,II}^{A-3}$, on obtient une imputation de son recensement pour l'année cible *A-2* en faisant :

$$\tilde{R}_{C,II}^{A-2} = R_{C,II}^{A-3} \times \frac{Adm_{II}^{A-2}}{Adm_{II}^{A-3}} = R_{C,II}^{A-3} \times \frac{\sum_{c \in II} Adm_c^{A-2}}{\sum_{c \in II} Adm_c^{A-3}},$$

où Adm_c^a est la valeur des sources administratives pour la commune *c* et l'année *a*.

En régime permanent, pour une «petite commune » enquêtée en *A-5* et *A* (voir la figure ci-dessous), on aura mesuré des variables sur les personnes (âge, sexe, activité, profession,...) et sur les logements (taille du ménage, nombre de pièces, mode d'occupation, confort...) aux deux moments.

	A-6		A-5		A-4		A-3		A-2		A-1		A	
Gr I		Adm		Adm	R	Adm		Adm	?	Adm		Adm		Adm
Gr II		Adm		Adm		Adm	R_{II}^{A-3}	Adm	$\tilde{R}_{C,II}^{A-2}$	Adm		Adm		Adm
Gr III		Adm		Adm		Adm		Adm	R_{III}^{A-2}	Adm		Adm		Adm
Gr IV	R	Adm		Adm		Adm		Adm	?	Adm	R	Adm		Adm
Gr V		Adm	R	Adm		Adm		Adm	?	Adm		Adm	R	
Total	5R	∑Adm	5R	∑Adm	5R	∑Adm	5R	∑Adm	5R	∑Adm	5R	∑Adm	5R	∑Adm

De plus, pour les groupes IV et V, l'estimation synthétique pour *A-2* pourrait profiter des informations recueillies durant la campagne de l'année *A-1* (respectivement *A*) ; en effet, il serait possible de calculer des facteurs d'ajustement par rapport au plus récent recensement et rétropoler sur la période intercensitaire. Par exemple, pour une commune *D* du groupe IV, on peut faire :

$$\Theta_1 = R_{D,IV}^{A-6} \times \frac{\sum_{c \in IV} Adm_c^{A-2}}{\sum_{c \in IV} Adm_c^{A-6}} \text{ et } \Theta_2 = R_{D,IV}^{A-1} \times \frac{\sum_{c \in IV} Adm_c^{A-2}}{\sum_{c \in IV} Adm_c^{A-1}}.$$

Il est à peu près certain que ces deux séries, extrapolations et rétropolations, ne coïncideront pas. Toutefois, il est souhaitable de publier une et une seule série d'estimations pour toute zone pour tout moment. Il apparaît naturel de produire une série «composite » dont les extrémités soient ancrées aux valeurs du recensement. La combinaison linéaire suivante peut jouer ce rôle en donnant plus d'important à la collecte la plus récente:

$$\tilde{R}_{D,IV}^{A-2} = 0.2 \times \Theta_1 + 0.8 \times \Theta_2.$$

De même, en donnant les définitions appropriées à Θ_1 et Θ_2, on ferait pour une commune *E* du groupe V :

$$\tilde{R}_{E,V}^{A-2} = 0.4 \times \Theta_1 + 0.6 \times \Theta_2.$$

Ces facteurs d'ajustement Θ devront être calculés pour des strates relativement fines de la population, des classes d'âge–sexe par exemple, de façon à préserver la plus grande souplesse démographique et géographique à l'ajustement des recensements. La qualité des fichiers administratifs et les disparités locales dicteront le niveau auquel l'ajustement peut être réalisé convenablement. On peut tenir un raisonnement analogue en grande commune si on remplace une « petite commune » par un « adresse ».

Finalement, quand toutes les communes de tous les groupes auront été imputées, il est improbable que l'estimation du total d'une variable d'intérêt obtenu du fichier imputé (estimations détaillées) ne corresponde

plus au total estimé à partir des seules observations (estimations globales publiées deux ans auparavant). Il est donc convenu que les estimations détaillées soient calées aux estimations globales. Le niveau de calage dépendra encore une fois des tendances locales et de la qualité des estimations globales.

3.3. Estimations de populations légales

La série des estimations des populations légales sont la troisième série tirée du recensement. Il s'agit des chiffres de population auxquels se réfèrent les textes de loi pour le financement des communes, le découpage électoral, la composition du conseil municipal, …

La « population totale » légale d'une commune comprend :

- les individus dont la résidence principale est située dans la commune,
- ceux qui résident dans un établissement ou un logement collectif situé sur le territoire de la commune,
- ceux qui résident dans un établissement ou un logement collectif dans situé dans une autre commune mais qui ont gardé un logement dans leur commune d'origine,
- les personnes qui vivent dans un logement collectif d'une autre commune pour leur travail ou dans une autre commune pour leurs études,
- et les populations administrativement rattachées à la commune (forains, mariniers, etc.).

On voit donc que ces populations ne peuvent être estimées qu'une fois l'ensemble du territoire couvert, c'est-à-dire qu'au moment des estimations détaillées.

3.4. Estimation de la variance d'échantillonnage

Il est prévu d'accompagner les estimations d'une mesure de leur qualité statistique. Les travaux portant sur cet aspect ont débuté à l'automne 2001 ; l'option privilégiée pour l'instant est le recours à des tables de référence, comme on le fait pour l'enquête canadienne sur la population active, par exemple. Les variances d'échantillonnage seront vraisemblablement obtenues par ré-échantillonnage sur la base de sondage.

3.5. Imprécision due à la synthèse

On a montré, à la section précédente, comment la production des estimations par synthèse utilise l'information amassée : d'abord une extrapolation pour un « vieux » recensement, pour deux groupes de rotation (disons I et II) ; ensuite l'utilisation directe des résultats du recensement pour un troisième groupe de rotation (disons III) ; enfin, la combinaison des extrapolations et rétropolations pour caler les deux derniers groupes (disons IV et V).

Cette synthèse peut être formalisée sous l'angle d'un modèle de non-réponse (Särndal, (1990)) : la campagne annuelle s'apparente alors à un sondage à 100% qui subirait 80% de non-réponse, laquelle est palliée par le recours à l'imputation par le ratio. Si l'échantillon complet est noté *s*, les répondants sont notés *r* et les non-répondants sont notés *s*-*r*, on peut écrire

$$y_{\bullet k} = \begin{cases} y_k \text{ si } k \in r \\ \hat{\beta}\, x_k \text{ si } k \in s - r \end{cases} \text{ avec } \hat{\beta} = \frac{\overline{y}_r}{\overline{x}_r}$$

C'est-à-dire que le modèle d'imputation est

$$\xi : \begin{cases} y_k = \beta\, x_k + \varepsilon_{k.} \\ E(\varepsilon_k) = 0 \\ V(\varepsilon_k) = \sigma^2 x_k \end{cases}$$

Avec un tel modèle d'imputation, sous sondage aléatoire simple,

$$\hat{Y}_\bullet = \frac{N}{n} \sum y_{\bullet k} = \frac{N}{n} \left\{ \sum_r y_k + \sum_{s-r} \hat{\beta} x_k \right\} = \ldots$$

$$= N \frac{\overline{y}_r}{\overline{x}_r} \overline{x}_s$$

L'incertitude autour de l'estimation avec imputation dépend des aléas de sondage et de la qualité du modèle d'imputation ξ :

$$\begin{array}{ccccc} (\hat{Y}_\bullet - Y) & = & (\hat{Y} - Y) & + & (\hat{Y}_\bullet - \hat{Y}) \\ \text{incertitude totale} & = & \text{incertitude du sondage} & + & \text{incertitude du modèle} \end{array}$$

Cela suppose que l'imputation se fait sans biais :

$$E_\xi E_s E_r (\hat{Y}_\bullet - Y) = 0$$

Donc,

$$\begin{aligned} V_{totale} &= E_\xi E_s E_r (\hat{Y}_\bullet - Y)^2 = ... \\ &= E_\xi E_s E_r (\hat{Y} - Y)^2 + E_\xi E_s E_r (\hat{Y}_\bullet - \hat{Y})^2 \\ &= E_\xi V_s + E_s E_r V_\xi \\ V_{totale} &= V_{échantillon} + V_{imputation} \end{aligned}$$

Pour de nombreux modèles d'imputation, l'utilisation de données imputées comme si elles avaient été observées dans le calcul de l'estimation de V_s même à une sous-estimation de $V_{\text{échantillon}}$. En espérance,

$$E_\xi \left(\hat{V}_s - \hat{V}_{\bullet s} \right) = V_{dif} \, .$$

Pour les estimateurs de ces variances, Särndal montre qu l'on obtient

$$\hat{V}_{sondage} = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \left\{ S_\bullet^2 + C_0 \hat{\sigma}^2 \right\}$$

$$\text{avec } C_0 \text{ près de } \left(1 - \frac{m}{n} \right) \bar{x}_{s-r} \text{ et } \hat{\sigma}^2 \text{ près de } \frac{\sum_r e_k^2}{\sum_r x_k} \text{ et}$$

$$\hat{V}_{imputation} = N^2 \left(\frac{1}{m} - \frac{1}{n} \right) A \bar{x}_s \hat{\sigma}^2 \ ,$$

avec $A = \frac{\bar{x}_{s-r}}{\bar{x}_r}$, qu'on peut comprendre comme un effet de sélection des répondants. On remarque que si x_k ≡1, alors, on n'impute pas et on obtient un sondage à deux phases de taille *m* parmi *n* et *n* parmi *N*. De plus, si *s* = *r*, $V_{totale} = V_{sondage}$.

Dans le modèle de Särndal, les x et y sont contemporains; à tout le moins, on aura observé certains des y. En reprenant la structure développée à la section précédente, on aurait :

Année *A-2*	
y_k	x_k
$y_{\bullet k}$	x_k

m répondants (Groupe II)
n-m imputations (autres groupes)

Dans l'application du RRP, tout n'est pas synchrone :

… A-4		A-3		A-2		A-1	A
Y_I^{A-4}	X_I^{A-4}		X_I^{A-3}		$X_I^{A—2}$		
	X_{II}^{A-4}	Y_{II}^{A-3}	X_{II}^{A-3}	$Y_{\bullet II}^{A-2}$	X_{II}^{A-2}		
	X_{III}^{A-4}		X_{III}^{A-3}	Y_{III}^{A-2}	X_{III}^{A-2}		
	X_{IV}^{A-4}		X_{IV}^{A-3}		X_{IV}^{A-2}	… …	
	X_V^{A-4}		X_V^{A-3}		X_V^{A-2}		… …

En effet, Y_{II}^{A-3}, X_{II}^{A-3} , $Y_{\bullet II}^{A-2}$, et X_{II}^{A-2} ne sont pas toutes mesurées ou observées la même année. En fait, quand on ne regarde que le seul groupe de rotation III par exemple, on a un échantillon de taille *n* en *A-2* et un échantillon identique en *A-3* entièrement non-répondant. En conséquence, certains paramètres de l'estimation de V_{totale} ne sont plus calculables.

En revanche, en regardant le problème pour un temps donné, on a bien un échantillon de taille *n* répondants et *4n* non-répondants. On pourrait approcher l'incertitude du processus d'imputation asynchrone (celui du recensement rénové) par celle du processus d'imputation synchrone (proche du modèle de Särndal).

Cette approche a été testée sur les petites communes de Rhône-Alpes, pour lesquelles les groupes de rotation, la Taxe d'Habitation (TH90) et le Recensement général de la population de 1990 (RGP90) sont disponibles.

4. TRAVAUX EN COURS

Les travaux de méthodologie autour de la rénovation du recensement sont loin d'être complétés. Au nombre des chantiers ouverts, notons

- l'établissement des règles de passage de seuil, les problèmes d'oscillation autour du seuil des 10,000 habitants et le calcul des populations légales;
- la sensibilité des bornes de strate en grande commune et leur robustesse dans le temps;
- la mise à jour et la maintenance des bases de sondage et des échantillons, en particulier, les ajustements éventuels suite aux passages de seuil et l'incorporation de nouveaux objets dans les groupes de rotation;
- l'imputation massive et la synthèse, tant les modèles que leur précision;
- l'estimation de la précision des estimateurs; et
- la collecte auprès des populations mobiles.

BIBLIOGRAPHIE

Bertrand, P., (2000), E*stimations 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), « Echantillonnage 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 JC. (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.

ONU (1990), *Principes et recommandations complémentaires concernant les recensements de la population et de l'habitat*, Etudes 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.