Techniques d'enquête, juin 1988
Vol. 14, n° 1, pp. 75–92
Statistique Canada

Test des opérations de redressement de 1986 dans le Central Los Angeles County

GREGG DIFFENDAL[1]

RÉSUMÉ

En vue du recensement décennal de 1990, le Census Bureau a étudié la possibilité de redresser les chiffres du recensement pour tenir compte du taux de sous-dénombrement estimé. À cette fin, il a exécuté un recensement d'essai dans un secteur à majorité hispanique du Central Los Angeles County afin de vérifier le calendrier et les caractéristiques des opérations de redressement réalisées au moyen d'une enquête post-censitaire (EP). Cet article vise à exposer les méthodes qui ont été utilisées pour produire des données de recensement qui tiennent compte de la population non dénombrée; on y trouve également les résultats de l'application de ces méthodes. Les méthodes utilisées pour le redressement des données de recensement comprenaient l'élaboration d'un plan de sondage, l'estimation "bivalente" et le calcul d'estimations régionales. Le plan de sondage prévoyait un échantillon d'îlots stratifiés selon l'origine ethnique. L'appariement s'est fait par ordinateur tandis que les opérations de contrôle et de résolution ont été exécutées manuellement. L'estimateur "bivalent", appelé aussi estimateur de Petersen ou saisie-resaisie, a servi à estimer la population. À cause de la nature des recensements, les chiffres du recensement ont été redressés avant de servir au calcul de l'estimateur bivalent. Avant de corriger les estimations régionales, on a ajusté un modèle de régression au facteur de redressement (estimateur bivalent divisé par le chiffre du recensement) afin de réduire les effets de la variabilité d'échantillonnage. Un estimateur synthétique a permis d'effectuer le redressement jusqu'au niveau de l'îlot. Les résultats de l'estimation bivalente sont présentés pour la région d'essai selon les trois principaux groupes ethniques (hispanique, asiatique, autre), le mode d'occupation, l'âge et le sexe. Enfin, nous présentons en bref les résultats du redressement des estimations régionales du recensement par îlot et nous en faisons l'analyse.

MOTS CLÉS: Sous-dénombrement dans le recensement; estimation bivalente; estimation synthétique; enquête post-censitaire.

1. INTRODUCTION

Depuis le premier recensement de la population des É.-U. en 1790, on a toujours eu de la difficulté à dénombrer toutes les personnes qui devaient l'être. Grâce aux progrès réalisés en démographie et en statistique, on a pu établir à partir de 1950 des estimations du taux de couverture du recensement. Ces estimations ont servi à évaluer les faiblesses du recensement et à définir les points qui devaient être améliorés en vue des recensements subséquents. Les estimations du taux de couverture du recensement ont connu une amélioration soutenue depuis la publication des premières estimations du genre en 1950. Une série d'estimations indique que le taux de sous-dénombrement pour l'ensemble du pays était de 4.4% en 1950, de 3.3% en 1960, de 2.8% en 1970 et de 1% en 1980. Malgré cette baisse soutenue du taux de sous-dénombrement, les estimations demeurent élevées pour certains groupes. Dans le cas des personnes de race noire, par exemple, le taux de sous-dénombrement se maintient à environ 5 points de pourcentage au-dessus de la moyenne nationale.

Les résultats indiquent aussi des taux de sous-dénombrement élevés pour d'autres groupes ethniques, notamment la population d'origine hispanique. De même, les taux de sous-

[1] Gregg Diffendal, Undercount Research Staff, Statistical Research Division, Bureau of the Census, Washington (D.C.) 20233, É.-U. Cet article est un compte rendu des recherches qui ont été faites par un membre du personnel du Census Bureau. Les opinions qui y sont exprimées sont celles de l'auteur et ne reflètent pas nécessairement la position du Census Bureau.

dénombrement sont plus élevés dans les grandes villes que dans les régions rurales et plus élevés chez les hommes que chez les femmes. Les membres du groupe d'âge 20-45 présentent aussi un fort taux de sous-dénombrement.

Depuis 1950, on mesure le taux de sous-dénombrement aux États-Unis au moyen d'une enquête post-censitaire (EP) et d'une analyse démographique. Le Census Bureau a laissé savoir que ces deux méthodes seraient les principaux moyens d'estimer le taux de sous-dénombrement pour le recensement de 1990. Une EP utilise un échantillon indépendant de personnes qui sont appariées aux enregistrements du recensement afin d'estimer la population totale. Marks (1978) et le U.S. Census Bureau (1979) décrivent les expériences qui ont déjà été faites au moyen d'une EP pour mesurer le taux de couverture du recensement. Par ailleurs, l'analyse démographique fait appel aux registres des naissances et des décès et à d'autres fichiers administratifs pour estimer la population totale des États-Unis. Fay et coll. (1988) présentent les estimations des taux de sous-dénombrement de 1980 qui ont été établies à l'aide de l'analyse démographique et du programme post-censitaire (PP).

En 1980, une analyse plus approfondie des chiffres du recensement a donné lieu à un certain nombre de poursuites judiciaires par lesquelles on exigeait un redressement de ces chiffres. Ces poursuites avaient été engagées pour des raisons comme la différence de taux de sous-dénombrement entre les noirs et les non-noirs, la création dans les années 1970 d'un mécanisme de partage des revenus qui liait directement des fonds publics aux chiffres de population et la baisse de population dans des villes et des États pour lesquels le taux de sous-dénombrement avait toujours été supérieur à la moyenne. Le U.S. Census Bureau s'est refusé au redressement des chiffres de 1980 en faisant valoir que le calcul du taux de sous-dénombrement était exposé aux erreurs et qu'un redressement des chiffres du recensement ne serait pas vraiment utile.

Le Census Bureau s'est néanmoins engagé après 1980 dans un programme de recherche visant à étudier diverses façons d'améliorer le calcul des taux de sous-dénombrement (Mulry et coll. 1981 et Hogan 1984). À ce sujet, Hogan (1984) a proposé une série de tests à effectuer concurremment avec les recensements d'essai. Cela a commencé par une EP à Tampa (Floride) en 1985 dans le but de tester et d'évaluer l'appariement par ordinateur. Ce test a confirmé la faisabilité de ce mode d'appariement (Jaro et Childers 1986). Des recensements d'essai et des EP ont aussi été réalisés en 1986 à Los Angeles et au Mississipi. Dans le premier cas, l'EP visait à tester le calendrier et les caractéristiques des opérations de redressement des chiffres du recensement. Dans le second cas, l'EP visait à évaluer les opérations d'une telle enquête dans une région d'essai rurale (Anolik 1988).

Une enquête précensitaire (EPR) a aussi été réalisée à Los Angeles en 1986 pour déterminer s'il était possible de réduire le délai des opérations de redressement en exécutant une partie du travail sur le terrain avant le recensement plutôt qu'après comme dans une EP (Wolfgang 1987). En 1987, on a réalisé une enquête précensitaire dans le Dakota du Nord afin d'évaluer les opérations d'une telle enquête dans des régions rurales; dans ce cas toutefois, on a procédé à des interviews sur place au lieu d'envoyer des questionnaires par la poste comme cela s'était fait dans les autres régions d'essai. Enfin, on est à préparer la répétition générale de 1988. Cette répétition générale permettra de tester toutes les opérations de recensement avant l'exécution du recensement décennal de 1990.

Cet article porte plus spécialement sur l'EP de 1986 pour le Central Los Angeles County, appelée Test des opérations de redressement (TOR) et réalisée concurremment avec le recensement d'essai. La région d'essai comprenait trois grands groupes ethniques: les personnes d'origine hispanique (qui constituaient environ 75% de la population totale), les personnes d'origine asiatique (environ 15% de la population totale) et les autres personnes, en majorité de race blanche (environ 10% de la population totale). L'EP a révélé un taux de sous-dénombrement estimé de 9%. Pour les trois principaux groupes ethniques de la région d'essai, les taux de sous-dénombrement estimés étaient de 9.8% pour les personnes d'origine hispanique, de 7.3% pour les personnes d'origine asiatique et de 6.2% pour les autres. Dans cet article, nous décrivons la manière dont nous estimons ces taux.

Dans la Section 2, nous exposons la méthode utilisée en 1986 pour estimer les taux de sous-dénombrement et indiquons comment ces estimations servent au redressement des chiffres du recensement. Dans la Section 3, nous analysons le calendrier des opérations prévu pour le TOR, y compris les opérations sur le terrain et l'appariement. La Section 4 contient un résumé des estimations des taux de sous-dénombrement pour les strates formées a posteriori et les îlots. Enfin dans la Section 5, nous faisons un résumé des principaux résultats de notre analyse et tirons quelques conclusions.

2. MÉTHODOLOGIE

2.1 Aperçu des échantillons utilisés dans l'estimation

Pour estimer la population, on a utilisé deux échantillons dans l'EP: l'échantillon P (pour population) et l'échantillon D (pour dénombrement). Le premier sert à mesurer les omissions du recensement tandis que le second sert à mesurer les enregistrements erronés du recensement.

L'échantillon P est un échantillon d'îlots qui s'accompagne d'un listage indépendant d'unités de logement et d'interviews sur place tandis que les données de l'échantillon D sont les enregistrements (chiffres) du recensement qui se rapportent aux îlots échantillonnés. L'échantillon P a servi à recueillir les données nécessaires à l'appariement et à l'estimation, y compris le domicile au jour du recensement. Pour les besoins du plan de sondage, il a fallu définir les éléments de l'échantillon P. L'échantillon P était constitué de toutes les personnes qui demeuraient aux adresses échantillonnées au moment de l'EP. Nous aurions pu choisir d'interviewer les personnes qui demeuraient à ces adresses le jour du recensement. Nous avons éliminé cette possibilité car tous les cas de déménagement impliquent une interview par personne interposée (c'est-à-dire une interview réalisée auprès d'une personne qui n'est pas membre du ménage). Selon la méthode que nous avons adoptée, des personnes ayant déménagé pouvaient se trouver à l'adresse échantillonnée et on pouvait les interviewer sans avoir affaire à un enquêté-substitut. Cependant, toutes les personnes qui ont quitté la région d'essai entre le jour du recensement et le jour de l'EP n'avaient aucune chance de faire partie de l'échantillon P. Par ailleurs, toutes les personnes de l'échantillon P qui demeuraient à l'extérieur de la région d'essai le jour du recensement étaient considérées comme exclues du champ de l'enquête.

Après l'interview, toutes les personnes de l'échantillon P ont été appariées à des enregistrements du recensement, l'appariement se faisant par ordinateur avec une vérification manuelle. Il a aussi fallu définir pour les besoins du plan de sondage l'étendue du secteur de recherche pour l'appariement. Selon l'EP, une personne de l'échantillon P était considérée comme appariée si elle avait été enregistrée lors du recensement n'importe où dans la région d'essai. On aurait pu définir un secteur de recherche moins étendu comme l'îlot servant à l'EP et les îlots avoisinants. Ainsi, il y a concordance pour une personne de l'échantillon P seulement lorsque si la personne correspondante dans le recensement se trouve dans ce secteur de recherche. À ce propos, l'EP de 1990 utilisera un secteur de recherche restreint pour l'appariement.

Tous les cas qui n'ont pu être appariés ont dû faire l'objet d'un suivi qui visait à recueillir des renseignements additionnels pour l'appariement. La tâche des personnes affectées au suivi de l'échantillon P a été sensiblement allégée du fait que toutes les questions nécessaires à l'appariement avaient été posées au moment de l'interview initiale. Par conséquent, seuls les cas où il manquait de l'information sur les caractéristiques personnelles, les cas où l'adresse de personnes ayant déménagé était incomplète et les cas d'appariement incertains ont dû faire l'objet d'un suivi. Les cas non appariés de l'échantillon P ont été jugés résolus et n'ont plus été renvoyés à l'équipe chargée du suivi. De nombreuses personnes de l'échantillon D ont pu être appariées à des personnes de l'échantillon P et ces cas ont été résolus sans devoir réaliser une autre interview. Tous les cas de l'échantillon D qui n'ont pu être résolus à l'aide de l'interview de l'échantillon P ont dû faire l'objet d'une interview de rappel qui visait à déterminer leur code de dénombrement. Les opérations sont analysées plus en détail à la section suivante. Les types d'enregistrements erronés que sert à mesurer l'échantillon D comprennent l'erreur de

géocodage, la répétition, l'information fausse, les personnes nées après le jour du recensement, les personnes décédées avant le jour du recensement et les cas non appariables. On dit qu'il y a erreur de géocodage lorsqu'un enregistrement du recensement existe à l'extérieur du secteur de recherche ou de la région d'essai. Les cas non appariables sont des enregistrements sans nom. Ils entraînent une surestimation du nombre d'enregistrements erronés mais sont traités comme ceux-ci dans le calcul de l'estimateur.

2.2 Estimation bivalente

Afin d'estimer la population totale, on utilise un estimateur bivalent qui réunit les données des échantillons P et D. Wolter (1986) décrit divers estimateurs bivalents et les hypothèses sur lesquelles ils reposent. L'estimateur bivalent utilisé dans le TOR est défini

$$\text{EB} = \frac{N_p(\text{REN-SUB-EE})}{\text{M}}, \tag{1}$$

où N_p est l'estimateur de la population totale de l'EP, REN est le chiffre du recensement non redressé, SUB est le nombre de substitutions de personnes dans le recensement, EE est l'estimateur du nombre d'enregistrements erronés et de personnes non appariables dans le recensement selon l'échantillon D et M est l'estimateur du nombre de personnes qui font partie à la fois de la population du recensement et de celle de l'EP. On parle de substitution de personne lorsqu'une personne quelconque a été enregistrée dans le recensement avec moins de deux caractéristiques démographiques. Afin de mieux illustrer quelques-unes des caractéristiques uniques de l'estimateur bivalent, nous exposons dans le tableau ci-dessous la classification des membres de la population.

Les effectifs du tableau 1 sont estimés à l'aide des composantes de l'estimateur bivalent: $\hat{N}_{11} = M$, $\hat{N}_{+1} = N_p$, $\hat{N}_{1+} = \text{REN-SUB-EE}$. La valeur de N_{22} est inobservable par définition mais on peut l'estimer en supposant que le recensement et l'échantillon P de l'EP sont indépendants l'un de l'autre. L'estimation de N_{22} est définie par l'équation suivante:

$$\hat{N}_{22} = \hat{N}_{12}\hat{N}_{21}/\hat{N}_{11}. \tag{2}$$

À l'aide des estimateurs définis ci-dessus, nous estimons la population totale par $\hat{N}_{++} = \text{EB}$.

À cause des problèmes que soulève l'appariement des données de recensement, il faut effectuer des opérations particulières pour ne pas surestimer la population. L'estimateur bivalent suppose que chaque personne est classée dans une seule case du tableau 1. Ainsi, au lieu d'utiliser simplement les chiffres du recensement, nous soustrayons de ces chiffres l'estimation du nombre d'enregistrements erronés pour obtenir une estimation du nombre **réel** de personnes dénombrées dans le recensement. En outre, l'estimateur bivalent suppose que chaque personne peut être désignée comme une concordance ou une non-concordance. Comme les enregistrements du recensement qui ne renferment pas assez d'information pour les besoins de l'appariement (par exemple, absence de nom ou moins de deux caractéristiques démographiques) ne

Tableau 1
Classification bivalente

		Population cible de l'échantillon P		
		Inclus	Exclus	Total
Recensement	Inclus	N_{11}	N_{12}	N_{1+}
Énumeration	Exclus	N_{21}	N_{22}	N_{2+}
	Total	N_{+1}	N_{+2}	N_{++}

Tableau 2
Strates d'échantillonnage et répartition des îlots échantillonnés

Strates d'échantillonnage	Nombre d'îlots échantillonnés
1. Îlots à majorité hispanique constitués de grands immeubles à logements multiples	8
2. Îlots à majorité hispanique constitués de petits immeubles à logements multiples	49
3. Îlots à majorité hispanique constitués de maisons unifamiliales	39
4. Îlots à majorité asiatique	35
5. Autres îlots	38
6. Îlots comptant deux unités de logement ou moins	21

peuvent être classés avec certitude parmi les concordances ou les non-concordances, nous avons aussi soustrait des chiffres du recensement le nombre de personnes non appariables. Toutes les personnes correspondantes de l'échantillon P ont été désignées comme des non-concordances et classées dans la case N_{21}.

2.3 Plan de sondage

Le plan de sondage prévoyait un échantillonnage stratifié où l'îlot servait d'unité d'échantillonnage. Deux catégories de données ont servi à stratifier la région d'essai: le nombre d'unités de logement par îlot, tiré du fichier d'adresses du recensement de 1986, et les résultats de l'application des données du recensement de 1980 sur l'origine ethnique aux unités géographiques du recensement de 1986. Comme cette opération ne pouvait s'effectuer qu'au niveau du secteur de recensement, qui peut contenir de un à six îlots, la répartition des groupes ethniques s'est faite à ce niveau. Tous les îlots d'un secteur de recensement ont été identifiés au même groupe ethnique et se sont trouvés par conséquent dans la même strate.

La région d'essai a été divisée en six strates d'échantillonnage comme l'indique le tableau 2.

Tous les îlots qui comprenaient des lieux particuliers (le plus souvent des foyers collectifs) ont été classés dans une strate d'échantillonnage distincte. Ces îlots étaient considérés comme exclus du champ de l'enquête et n'ont pas été échantillonnés. Les petits îlots ont été classés dans une strate distincte afin de réduire la variance d'échantillonnage. Tous les îlots des secteurs de recensement qui comptaient au moins 18% d'Asiatiques ont servi à former la strate à majorité asiatique. Tous les îlots à majorité non asiatique des secteurs de recensement qui comptaient au moins 40% de personnes d'origine hispanique ont servi à former les trois strates à majorité hispanique. Les îlots qui n'étaient pas inclus dans l'une ou l'autre des strates précédentes ont servi à former la strate réservée aux personnes d'autres origines.

Les données de 1986 sur le logement renfermaient aussi de l'information sur les maisons unifamiliales et les immeubles à logements multiples. Ces données ont servi à partager les strates à majorité hispanique en trois catégories: maisons unifamiliales, petits immeubles à logements multiples et grands immeubles à logements multiples. La strate de la troisième catégorie était composée des îlots à majorité hispanique où au moins la moitié des unités de logement se trouvaient dans des immeubles de dix logements ou plus. La strate de la première catégorie était formée des îlots à majorité hispanique qui comptaient plus de 50% de maisons unifamiliales. Enfin, la strate de la seconde catégorie comprenait tous les autres îlots à majorité hispanique.

Des îlots ont été prélevés dans chacune des strates au moyen d'un échantillonnage systématique avec probabilités égales. L'échantillon était composé de 190 îlots renfermant environ

6000 unités de logement. Le tableau 2 donne la répartition des îlots échantillonnés selon les strates. Les grands îlots qui comptent 70 unités de logement ou plus ont fait l'objet d'un sous-échantillonnage afin de réduire la tâche des intervieweurs. Le sous-échantillonnage consistait à diviser l'îlot en grappes de 35 à 50 unités de logement à l'aide de tranches d'adresses ou de côtés d'îlot. On choisissait une grappe au hasard dans le but d'interviewer des membres de l'échantillon P. Quant à l'échantillon D, il comprenait toutes les personnes qui faisaient partie de cette grappe le jour du recensement.

2.4 Stratification a posteriori

L'estimateur bivalent est biaisé et ce biais peut être élevé s'il y a de fortes différences de taux de sous-dénombrement entre les sous-groupes de la population (Wolter 1986). Pour réduire au minimum ce biais, on a stratifié a posteriori la région d'essai selon des critères précis de manière à ce que les taux de sous-dénombrement soient comparables dans toutes les strates. On a ensuite calculé des estimations bivalentes dans chaque strate ainsi formée.

Les strates ont été choisies par suite d'un examen de la structure de la région d'essai et d'une analyse des données de l'EP de 1980. On a observé que l'origine ethnique était la variable qui influait le plus sur le taux de sous-dénombrement. On a donc défini trois groupes ethniques: les personnes d'origine hispanique, les personnes d'origine asiatique et les autres. Il n'a pas été possible de créer une strate pour les personnes de race noire puisque la région d'essai comptait très peu de noirs. Le fait d'être locataire constituait une variable explicative importante dans notre étude précédente (Isaki et coll. 1987). Par conséquent, le mode d'occupation a aussi été considéré dans la construction des strates. Par ailleurs, comme on croyait que les personnes d'origine hispanique vivant dans des îlots qui comptaient moins de 50% de personnes d'origine hispanique (appelés îlots à minorité hispanique) n'avaient pas le même taux de sous-dénombrement que les autres personnes de même origine, on les a incluses dans une strate à

Tableau 3

Catégories "origine ethnique — mode d'occupation" utilisées dans la stratification a posteriori, y compris la taille des échantillons

Catégories "origine ethnique — mode d'occupation"	Échantillon P	Échantillon D
Locataires d'origine hispanique vivant dans des îlots à majorité hispanique	8,182	8,739
Propriétaires d'origine hispanique vivant dans des îlots à majorité hispanique	5,688	5,867
Personnes d'origine hispanique vivant dans des îlots à minorité hispanique	896	1,005
Locataires d'origine asiatique	666	911
Propriétaires d'origine asiatique	1,144	1,230
Autres locataires	1,135	1,316
Autres propriétaires	1,841	1,908
Total	19,552	20,976

part. Le tableau 3 donne les sept catégories "origine ethnique — mode d'occupation" qui ont été subdivisées selon l'âge (0-14, 15-29, 30-44, 45-64, 65 +) et le sexe pour former les 70 strates qui ont servi à l'estimation.

Le tableau 3 indique également la taille des échantillons P et D. La taille de l'échantillon P est moins élevée que celle de l'échantillon D à cause notamment des membres de l'échantillon P qui sont venus s'établir dans la région d'essai après le recensement et qui sont exclus par conséquent du champ de l'enquête.

2.5 Traitement des données manquantes

Le calcul des estimations bivalentes exige un fichier de données complet. Or, comme dans le cas de n'importe quelle enquête par sondage, le test de 1986 n'a pas permis de recueillir toutes les données voulues. Schenker (1988) décrit les méthodes utilisées pour résoudre le problème des données manquantes et analyse notamment les effets de diverses hypothèses concernant les données manquantes sur les estimations bivalentes. Afin d'avoir une idée d'ensemble de la question, nous décrivons brièvement ici ces méthodes.

Dans le test de 1986, il manquait des données sur les caractéristiques des personnes et du ménage, le code d'appariement (apparié/non-apparié) pour les personnes de l'échantillon P et le code de dénombrement (exact/erroné) pour les personnes de l'échantillon D. Pour les cas de non-interview de l'échantillon P, on a eu recours à un redressement par pondération. L'imputation des caractéristiques manquantes s'est faite à l'aide d'une méthode "hot-deck".En ce qui concerne les codes d'appariement, on s'est servi d'un modèle de régression logistique pour estimer la probabilité d'appariement. Au lieu d'attribuer un code d'appariement définitif à chaque cas non résolu, on s'est servi des probabilités estimées dans le calcul des estimations bivalentes. On a procédé de façon analogue pour les codes de dénombrement manquants de l'échantillon D.

2.6 Estimation relative aux petites régions

Afin que les utilisateurs disposent de données redressées à tous les niveaux d'agrégation, on calcule des estimations du taux de sous-dénombrement jusqu'au niveau de l'îlot, qui est la plus petite unité géographique. Avant d'effectuer ces calculs toutefois, on utilise un modèle de régression pour "lisser" les effets de l'erreur d'échantillonnage. Dans ce modèle, des facteurs de redressement servent de variable dépendante. Un facteur de redressement est défini comme le quotient de l'estimateur bivalent par le chiffre du recensement:

$$Y = \text{EB/REN}, \tag{3}$$

où REN et EB sont définis comme auparavant.

Le modèle de régression s'exprime par le formule suivante:

$$Y_i = B_o + B_1 X_{i1} + \ldots + B_p X_{ip} + S_i + E_i, \tag{4}$$

où Y_i est le facteur de redressement pour la i-ième strate formée a postériori ($i = 1, \ldots, 70$), X_{ij} = est la variable indépendante ($j = 1, \ldots, p$), B_j est le coefficient de régression à estimer, $S_i$est l'erreur d'échantillonnage du facteur de redressement, E_i est l'erreur du modèle et les valeurs de S_i et de E_i sont indépendantes et distribuées selon une loi normale avec une moyenne nulle et des variances égales à σ_i^2et à ϵ^2 respectivement. Les valeurs d'ϵ^2 et de B_j sont estimées à l'aide des méthodes du maximum de vraisemblance (Ericksen et Kadane 1985). Les valeurs de σ_i^2 sont estimées directement à l'aide de l'échantillon. On fait la moyenne du facteur de

redressement fondé sur l'échantillon et du facteur de redressement fondé sur le modèle pour obtenir le facteur de redressement prévu

$$AF_i = \left(Y_i/\sigma_i^2 + \sum_j X_{ij}\hat{B}_j/\epsilon^2\right)\left(\sigma_i^{-2} + \epsilon^{-2}\right)^{-1}, \tag{5}$$

qui sert à corriger les données de recensement relatives à l'îlot. La variance de AF_i peut être déduite des résultats publiés dans Freedman et Navidi (1986).

L'estimation synthétique a servi à reproduire le redressement effectué dans chaque strate au niveau de l'îlot. L'estimateur synthétique est défini

$$ADJ_{iJ} = AF_i \times CEBN_{ij}, \tag{6}$$

où i et j désignent respectivement la strate formée a posteriori et l'îlot et ADJ est le chiffre de population redressé pour l'îlot.

Le chiffre de population redressé ADJ_{ij} est en règle générale un nombre fractionnaire. Or, le recensement sert à dénombrer des unités. Par conséquent, si nous voulons établir une compatibilité entre les opérations de redressement et le recensement, il est nécessaire de convertir les nombres fractionnaires en nombres entiers. Cette conversion (appelée arrondissement contrôlé) permet d'arrondir vers le haut ou vers le bas tous les nombres fractionnaires (Causey et coll. 1985).

Après avoir arrondi les estimations redressées des îlots, on a produit des chiffres selon l'âge, l'origine ethnique et le sexe qui indiquaient le nombre de personnes qu'il fallait ajouter ou retrancher dans chaque îlot. Dans le cas d'un sous-dénombrement, on prélevait aléatoirement dans l'îlot un enregistrement du recensement qui présentait la même série de caractéristiques que la personne oubliée et on en tirait un nouvel enregistrement identique. On a utilisé une catégorie n'ayant pas trait aux ménages pour ajouter des personnes dans les listes du recensement de manière qu'il ne soit pas nécessaire de tenir compte des rapports entre les membres du ménage ou de créer de nouveaux ménages. Zaslavasky (1988) présente une autre méthode, fondée sur la pondération, pour ajouter des personnes et des ménages dans les îlots de recensement. Dans le cas d'un surdénombrement, on signalerait les enregistrements qui présentent les caractéristiques voulues et on les incluerait dans les totalisations corrigées du recensement.

3. OPÉRATIONS ET CALENDRIER DES OPÉRATIONS

Le recensement d'essai avait principalement pour but d'analyser le calendrier et les caractéristiques des opérations de redressement des données du recensement. Les EP qu'avait réalisées le Census Bureau dans les années antérieures s'étaient étendues sur au moins deux ans. Par exemple, l'EP de 1980 a produit des estimations du taux de sous-dénombrement à l'automne de 1981 puis une série d'estimations finales au début de 1982.

Le tableau 4 énumère les principales opérations du recensement et de l'EP ainsi que les dates qui marquent le début et la fin de chacune de ces opérations. On remarquera qu'il s'écoule un certain temps entre la fin d'une opération et le début d'une autre car les opérations du recensement et de l'EP ne sont pas toutes indiquées. Certaines périodes se chevauchent car les opérations correspondantes se déroulaient simultanément. On a amorcé l'EP une fois que les principales opérations régionales touchant le recensement ont été terminées, cela ayant pour but d'éviter que le personnel des bureaux régionaux n'ait à exécuter simultanément les opérations des deux enquêtes.

Tableau 4
Calendrier des opérations pour le TOR de 1986

Opération	Début	Fin
Jour du recensement	16 mars	16 mars
Suivi des cas de non-réponse	9 avril	8 mai
Saisie des noms du recensement	23 mai	10 juin
Fichier du recensement pour l'appariement	8 août	15 août
Liste d'adresses pour l'EP	17 juin	21 juin
Sous-échantillonnage pour l'EP	25 juin	1 juillet
Période d'interview pour l'EP	25 juin	8 août
Saisie des données de l'EP	21 juillet	19 août
Appariement informatisé	28 août	9 septembre
Appariement informatisé (deuxième phase)	9 septembre	3 octobre
Appariement manuel	15 septembre	31 octobre
Suivi sur place	23 septembre	6 novembre
Appariement des résultats du suivi	29 septembre	6 novembre
Saisie des résultats de l'appariement	21 octobre	10 novembre
Préparation des fichiers des échantillons P et D	11 novembre	2 janvier
Imputation	5 janvier	11 janvier
Version finale du fichier du recensement	—	5 janvier
Estimation relative aux strates formées à posteriori	12 janvier	11 février
Estimation relative aux petites régions	12 février	22 février

Le recensement a été réalisé à l'aide d'un questionnaire qui a été posté à chaque unité de logement connue et qu'un membre du ménage devait remplir le jour du recensement (16 mars). Les ménages qui négligeaient de retourner leur questionnaire recevaient la visite d'un agent recenseur qui procédait à une interview sur place. C'est ce qu'on appelle le suivi des cas de non-réponse. Une fois remplis, les questionnaires étaient acheminés au service de traitement pour la saisie des données, qui comprenaient en l'occurrence tous les noms du recensement.

La première étape de l'EP consistait à dresser une liste indépendante de toutes les adresses figurant dans les îlots échantillonnés. On a confronté la liste en question avec une liste administrative pour s'assurer de son exactitude et de son intégralité. Cette mesure de contrôle de la qualité a permis de constater que 127 îlots (67%) ne requéraient aucune modification à la liste d'adresses. Les 63 autres îlots (33%) ont dû faire l'objet de modifications par l'intermédiaire du contrôle de qualité puis les adresses correspondantes ont été relistées. Cette opération a permis d'ajouter des adresses dans 37 îlots, d'en corriger dans 39 îlots et d'en soustraire dans 9 îlots. (Comme certains îlots ont fait l'objet de plus d'un genre de modifications, la somme des nombres ci-dessus ne correspond pas au nombre total d'îlots qui ont dû faire l'objet de modifications.)La mesure de contrôle de la qualité n'a entraîné que des corrections mineures pour les autres adresses. À l'issue de cette opération de contrôle, tous les îlots comptant 70 unités de logement ou plus ont été divisés par sous-échantillonnage à l'aide des côtés d'îlots ou des tranches d'adresses.

L'interview de l'EP a été réalisée sur place. Elle visait à recueillir des renseignements sur les caractéristiques démographiques de toutes les personnes qui demeuraient à l'endroit visité le jour de l'enquête. L'interview comportait des questions particulières comme celles portant sur le domicile le jour du recensement, l'adresse postale, d'autres lieux d'habitation comme les foyers d'étudiants, et les autres personnes qui pouvaient avoir demeuré à l'endroit visité le jour du recensement. Une mesure de contrôle de la qualité du questionnaire de l'EP a permis de vérifier la liste des noms. Quatre-vingt-seize pour cent des questionnaires vérifiés ont passé le test. Pour ce qui est des autres (4%), on a dû procéder à une nouvelle interview et corriger les questionnaires en conséquence.

À la fin de la période d'interview, 5,714 (93.2%) unités de logement avaient fait l'objet d'une interview complète avec un membre du ménage, 193 (3.1%) avaient été classées parmi les logements inoccupés et 189 (3.1%) avaient fait l'objet d'une interview complète avec une personne qui n'était pas membre du ménage (par exemple, un voisin). Seulement 32 (0.5%) unités de logement ont été classées parmi les cas de non-interview. Ce taux extrêmement faible s'explique par le fait que l'on avait alloué cinq semaines pour les interviews.

Une fois remplis, les questionnaires de l'EP ont été préparés en vue de l'appariement informatisé avec le fichier du recensement. L'appariement informatisé a été divisé en deux phases: la première visait à apparier les données de l'EP avec celles de l'échantillon D et la seconde visait à apparier tous les cas de l'échantillon P qui n'avaient pu l'être dans la première phase avec les autres données du recensement. La seconde phase de l'appariement informatisé servait aussi à apparier les personnes ayant déménagé entre le jour du recensement et le jour de l'interview pour l'EP avec les erreurs de géocodage (c.-à-d., lorsque l'unité de logement est associée au mauvais îlot). La première phase de l'appariement informatisé a permis de déterminer une concordance pour 14,700 cas (73.5%) de l'échantillon P et une concordance probable pour 2,550 autres cas (12.0%). La seconde phase de l'appariement informatisé a permis de déterminer une concordance pour 130 personnes (0.7%) et une concordance probable pour 570 autres personnes (2.9%). Étant donné le faible pourcentage de cas de l'échantillon P pour lesquels la seconde phase de l'appariement a permis de déterminer une concordance, nous en avons conclu que le géocodage dans Los Angeles comportait peu d'erreurs.

L'appariement manuel a servi à revoir les résultats de l'appariement informatisé. Il a servi aussi à identifier les cas qui n'avaient pu être appariés à cause d'un manque de données (pour ces cas, il faut recourir à l'imputation). Enfin, c'est durant l'appariement manuel qu'ont été rédigés les questionnaires de suivi pour les cas non résolus des échantillons P et D.

Le suivi sur place visait 1,551 unités de logement, dont 1,511 (97.4%) avaient fait l'objet d'une interview complète. Cette opération a été suivie d'un appariement final. Les résultats définitifs de l'échantillon P indiquent que 17,018 (85.2%) personnes ont pu être appariées à un enregistrement du recensement tandis que 2,373 (11.9%) n'ont pu l'être. En outre, 426 (2.1%) personnes ont été considérées comme exclues du champ de l'enquête (il s'agissait pour la plupart de personnes qui demeuraient à l'extérieur de la région d'essai le jour du recensement) et 161 (0.8%) personnes représentaient des cas non résolus (auxquels on a imputé ultérieurement un code d'appariement). Les résultats définitifs de l'échantillon D indiquent que 19,637 (93.6%) personnes ont été enregistrées correctement dans le recensement tandis que 360 (1.7%) l'ont été incorrectement. De plus, 976 (4.7%) personnes représentaient des cas non résolus auxquels on a imputé un code de dénombrement.

Après l'appariement final, toutes les données manquantes, y compris les codes d'appariement pour l'échantillon P et les codes de dénombrement pour l'échantillon D, ont été imputées. Les résultats ont servi à calculer les estimations bivalentes. Les estimations ont été lissées puis reproduites au niveau de l'îlot pour créer un fichier du recensement redressé. La réduction du délai de production des estimations du taux de sous-dénombrement a été surtout attribuable aux opérations d'appariement. Pour le TOR, l'appariement informatisé et l'appariement manuel se sont étendus sur environ trois mois tandis que les opérations d'appariement pour l'EP de 1980 s'étaient déroulées sur plus d'un an. Une meilleure planification des opérations ainsi qu'une facilité accrue pour consulter les documents du recensement sont deux facteurs qui ont aussi contribué à réduire les délais de production.

4. ESTIMATIONS

4.1 Estimations relatives aux strates formées a posteriori

Cette sous-section présente les estimations du taux de sous-dénombrement pour divers groupes de strates formées a posteriori. Le tableau 5 donne le taux de sous-dénombrement (100(1-REN/EB)), le pourcentage de cas non appariés (100(1-M/N_p)), le pourcentage de cas

Tableau 5

Taux de sous-dénombrement (en pourcentage) et composantes de l'estimateur bivalent pour les strates formées a posteriori

Strates formées a posteriori	Taux de sous-dénombrement[a]	Pourcentage de cas de l'échantillon P non appariés	Pourcentage de cas de l'échantillon D enregistrés incorrectement[b]	Pourcentage de cas ayant fait l'objet d'une subtitution dans le recensement
Locataires d'origine hispanique vivant dans des îlots à majorité hispanique	13.7	17.1	2.6	1.7
Propriétaires d'origine hispanique vivant dans des îlots à majorité hispanique	5.5	8.1	1.2	1.5
Personnes d'origine hispanique vivant dans des îlots à minorité hispanique	7.5	9.7	1.4	1.3
Locataires d'origine asiatique	11.1	13.4	2.1	1.2
Propriétaires d'origine asiatique	4.6	6.8	1.2	1.5
Autres locataires	9.9	12.9	2.4	1.7
Autres propriétaires	3.8	5.8	1.3	0.9
0-14	8.8	11.9	2.2	1.6
15-29	13.6	16.2	2.1	1.6
30-44	8.6	10.8	1.4	1.4
45-64	4.5	6.6	1.3	1.4
65+	3.3	5.9	1.7	1.3
Homme	9.7	12.1	1.7	1.5
Femme	8.3	10.8	1.9	1.5
Total	9.0	11.4	1.8	1.5

[a] La sommation de toutes les estimations est étendue à toutes les autres catégories.
[b] Les cas enregistrés incorrectement comprennent les cas non appariables qui n'ont pas fait l'objet d'une substitution.

enregistrés incorrectement (100(EE/REN)) et le pourcentage de cas ayant fait l'objet d'une substitution (100(SUB/REN)).

Une caractéristique de l'estimateur bivalent veut que la somme des estimations pour plusieurs catégories n'égale pas l'estimation directe de la somme pour l'ensemble de ces catégories. Ainsi, pour assurer la cohérence des estimations indiquées dans le tableau 5, on a étendu la sommation de toutes les estimations aux autres catégories pertinentes.

Si nous examinons les taux de sous-dénombrement indiqués dans le tableau 5 pour les catégories "origine ethnique-mode d'occupation", nous en concluons que le mode d'occupation est une bonne variable de stratification, les locataires ayant un taux de sous-dénombrement plus élevé que les propriétaires; l'origine ethnique influe aussi sur le taux de sous-dénombrement, les personnes d'origine hispanique ayant un taux plus élevé que les personnes d'origine

asiatique, qui à leur tour ont un taux plus élevé que les autres. En ce qui concerne les enregistrements erronés, le pourcentage de cas est plus élevé pour les locataires que pour les propriétaires alors que l'origine ethnique n'est pas une variable déterminante sur ce plan. Pour ce qui a trait aux cas ayant fait l'objet d'une substitution, le pourcentage est plus élevé pour les locataires d'origine hispanique et les autres locataires que pour les propriétaires d'origine hispanique et les autres propriétaires. En revanche, on observe la situation inverse dans le cas des Asiatiques.

Si nous examinons maintenant les taux de sous-dénombrement selon l'âge et le sexe, nous constatons que le groupe d'âge 15-29 a le taux le plus élevé et que les hommes ont un taux plus élevé que les femmes. Les groupes d'âge 0-14 et 30-44 présentent des taux de sous-dénombrement comparables, lesquels sont légèrement au-dessous de la moyenne pour la région d'essai. Les groupes d'âge 45-64 et 65 + présentent aussi des taux comparables, qui sont bien inférieurs à ceux des autres groupes d'âge. Au point de vue de la distribution, ces résultats correspondent assez bien à ceux observés antérieurement.

Le pourcentage de cas enregistrés incorrectement est le plus élevé pour les deux premiers groupes d'âge (0-14 et 15-29). Les deux groupes suivants (30-44 et 45-64) présentent les pourcentages les plus faibles. Curieusement, le groupe d'âge 65 + présente un pourcentage qui se situe entre les pourcentages des groupes précités. Le pourcentage de cas enregistrés incorrectement varie peu d'un sexe à l'autre. Enfin, pour ce qui a trait aux cas ayant fait l'objet d'une substitution, les différences de pourcentage entre les groupes d'âge et entre les sexes sont faibles.

4.2 Estimations relatives aux petites régions

Comme il a été mentionné précédemment, avant d'opérer un redressement des données au niveau de l'îlot, nous avons ajusté un modèle de régression dans le but de "lisser" les données et de réduire les effets de la variabilité d'échantillonnage. Le modèle de régression a été ajusté aux 70 facteurs de redressement qui avaient été définis par les strates formées a posteriori. Ce genre de modèle sert à déterminer un mode de calcul du taux de sous-dénombrement uniforme pour les régions ayant les mêmes caractéristiques. Dans un deuxième temps, on rajuste les facteurs de redressement fondés sur l'échantillon de manière à les rendre comparables aux estimations du modèle. Ce procédé s'inspire de l'estimateur de James-Stein et des estimateurs empiriques de Bayes. Les variables indépendantes qui pouvaient être utilisées en l'occurrence dans le modèle étaient des variables indicatrices pour les catégories "origine ethnique-mode d'occupation", les groupes d'âge et les sexes. Il n'était pas permis d'inclure des termes d'interaction dans le modèle. Le modèle qui pouvait être ajusté aux données et qui présentait des coefficients significatifs (selon un modèle de régression non pondéré) était le suivant:

$$\hat{Y} = 1.038 + .090(\text{HR}) + .044(\text{AR}) + .013(\text{OR}) + .058(\text{A15-29}) - .009(\text{A45-64})$$

où $\hat{Y}$ = facteur de redressement axé sur le modèle

LH = 1 si locataire d'origine hispanique vivant dans un îlot à majorité hispanique
= 0 dans les autres cas

LAS = 1 si locataire d'origine asiatique vivant dans un îlot quelconque
= 0 dans les autres cas

AL = 1 si autre locataire vivant dans un îlot quelconque
= 0 dans les autres cas

A15-29 = 1 si groupe d'âge 15-29
= 0 dans le cas contraire

A45-64 = 1 si groupe d'âge 45-64
= 0 dans le cas contraire.

Tableau 6

Résultats du lissage des facteurs de redressement du TOR

Strates formées a posteriori	Sexe/Âge	Estimé		Prévu	
		Facteur de redressement (Y)	Erreur type	Facteur de redressement (AF)	Erreur type
LH dans IH	H' 0-14	1.131	0.020	1.130	0.016
LH dans IH	H' 15-29	1.247	0.030	1.211	0.021
LH dans IH	H' 30-44	1.165	0.029	1.144	0.020
LH dans IH	H' 45-64	1.099	0.043	1.114	0.024
LH dans IH	H' 65+	1.055	0.044	1.110	0.023
LH dans IH	F 0-14	1.124	0.023	1.126	0.018
LH dans IH	F 15-29	1.234	0.032	1.203	0.022
LH dans IH	F 30-44	1.084	0.017	1.098	0.015
LH dans IH	F 45-64	1.125	0.040	1.121	0.024
LH dans IH	F 65+	1.099	0.045	1.122	0.024
PH dans IH	H' 0-14	1.056	0.018	1.050	0.015
PH dans IH	H' 15-29	1.078	0.018	1.084	0.015
PH dans IH	H' 30-44	1.087	0.016	1.072	0.014
PH dans IH	H' 45-64	1.031	0.012	1.031	0.011
PH dans IH	H' 65+	1.073	0.028	1.054	0.019
PH dans IH	F 0-14	1.059	0.020	1.051	0.016
PH dans IH	F 15-29	1.088	0.016	1.090	0.014
PH dans IH	F 30-44	1.033	0.012	1.034	0.011
PH dans IH	F 45-64	1.020	0.012	1.022	0.011
PH dans IH	F 65+	1.033	0.019	1.035	0.015
H dans INH	H' 0-14	1.105	0.052	1.051	0.023
H dans INH	H' 15-29	1.154	0.054	1.106	0.025
H dans INH	H' 30-44	1.131	0.065	1.050	0.024
H dans INH	H' 45-64	1.063	0.050	1.036	0.023
H dans INH	H' 65+	0.991	0.000	0.991	0.000
H dans INH	F 0-14	1.137	0.047	1.059	0.023
H dans INH	F 15-29	1.033	0.022	1.060	0.017
H dans INH	F 30-44	1.079	0.037	1.051	0.021
H dans INH	F 45-64	1.033	0.028	1.031	0.019
H dans INH	F 65+	0.947	0.040	1.013	0.022
LAS dans tous I	H' 0-14	1.059	0.041	1.076	0.026
LAS dans tous I	H' 15-29	1.127	0.044	1.137	0.028
LAS dans tous I	H' 30-44	1.195	0.077	1.093	0.031
LAS dans tous I	H' 45-64	1.004	0.057	1.063	0.030
LAS dans tous I	H' 65+	0.982	0.000	0.982	0.000
LAS dans tous I	F 0-14	1.067	0.047	1.079	0.028
LAS dans tous I	F 15-29	1.215	0.055	1.153	0.029
LAS dans tous I	F 30-44	1.173	0.105	1.087	0.032
LAS dans tous I	F 45-64	1.012	0.061	1.065	0.030
LAS dans tous I	F 65+	1.212	0.127	1.087	0.032

Tableau 6

Résultats du lissage des facteurs de redressement du TOR — Fin

Strates formées a posteriori	Sexe/Âge	Estimé		Prévu	
		Facteur de redressement (Y)	Erreur type	Facteur de redressement (AF)	Erreur type
PAS dans tous I	H' 0-14	1.045	0.030	1.041	0.019
PAS dans tous I	H' 15-29	1.059	0.038	1.085	0.022
PAS dans tous I	H' 30-44	1.091	0.040	1.053	0.022
PAS dans tous I	H' 45-64	1.035	0.020	1.033	0.016
PAS dans tous I	H' 65+	1.031	0.051	1.037	0.023
PAS dans tous I	F 0-14	1.040	0.041	1.039	0.022
PAS dans tous I	F 15-29	1.052	0.046	1.086	0.024
PAS dans tous I	F 30-44	1.035	0.036	1.037	0.021
PAS dans tous I	F 45-64	1.038	0.019	1.035	0.015
PAS dans tous I	F 65+	1.051	0.045	1.041	0.022
AL dans tous I	H' 0-14	1.037	0.059	1.049	0.027
AL dans tous I	H' 15-29	1.252	0.114	1.115	0.031
AL dans tous I	H' 30-44	1.144	0.066	1.062	0.028
AL dans tous I	H' 45-64	1.055	0.031	1.047	0.022
AL dans tous I	H' 65+	1.068	0.056	1.054	0.027
AL dans tous I	F 0-14	1.148	0.062	1.064	0.027
AL dans tous I	F 15-29	1.126	0.054	1.112	0.028
AL dans tous I	F 30-44	1.134	0.057	1.064	0.027
AL dans tous I	F 45-64	1.068	0.041	1.049	0.025
AL dans tous I	F 65+	0.948	0.021	0.992	0.018
AP dans tous I	H' 0-14	1.044	0.037	1.040	0.021
AP dans tous I	H' 15-29	1.148	0.064	1.103	0.025
AP dans tous I	H' 30-44	1.006	0.048	1.032	0.023
AP dans tous I	H' 45-64	1.036	0.017	1.034	0.014
AP dans tous I	H' 65+	1.017	0.019	1.025	0.016
AP dans tous I	F 0-14	1.159	0.068	1.052	0.024
AP dans tous I	F 15-29	1.081	0.042	1.092	0.023
AP dans tous I	F 30-44	0.997	0.017	1.011	0.014
AP dans tous I	F 45-64	1.025	0.012	1.026	0.011
AP dans tous I	F 65+	0.997	0.012	1.004	0.011

Nota: H: hispanique, L: locataire, I: îlot, H': Homme, F: femme, P: propriétaire, A: autre, NH: non hispanique, AS: asiatique.

(Exemple: LH dans IH: locataire d'origine hispanique vivant dans un îlot à majorité hispanique)

Le modèle de régression indique des taux de sous-dénombrement plus élevés pour les locataires que pour les propriétaires, peu importe l'origine ethnique. De plus, le groupe d'âge 15-29 présente un taux de sous-dénombrement beaucoup plus élevé que les autres groupes. Inversement, le groupe d'âge 45-64 est celui qui présente le taux le plus faible. La variable "sexe" n'étant pas statistiquement significative, elle n'a pas été incluse dans le modèle. Deux facteurs de redressement (hommes d'origine hispanique âgés de 65 et plus et vivant dans des îlots non hispaniques et locataires asiatiques de sexe masculin âgés de 65 et plus) avaient une variance estimée nulle et n'ont donc pas été inclus dans le modèle. Pour ces deux facteurs de redressement, le facteur fondé sur l'échantillon a tenu lieu de facteur prévu.

Le tableau 6 renferme les facteurs de redressement fondés sur l'échantillon (facteurs estimés) et les facteurs de redressement prévus pour les 70 strates formées a posteriori. En règle générale, les facteurs de redressement prévus sont inférieurs aux facteurs estimés les plus élevés mais supérieurs aux facteurs estimés les plus faibles. Les facteurs de redressement prévus ont une moins grande variabilité que les facteurs de redressement fondés sur l'échantillon. Les effets du modèle de régression sont les plus visibles dans le cas des locataires asiatiques de sexe féminin âgées de 65 et plus. Le facteur de redressement prévu pour cette strate est de 1.087 comparativement à un facteur de redressement estimé de 1.212. Dans ce cas, le facteur de redressement prévu est plus conforme à la représentation que l'on se fait du taux de sous-dénombrement pour les personnes de 65 ans et plus, c'est-à-dire un taux moins élevé que pour les autres groupes d'âge.

On a multiplié les chiffres du recensement par les facteurs de redressement prévus pour les 2 405 îlots de la région d'essai. On a ensuite arrondi les chiffres redressés pour obtenir des nombres entiers. Malgré que trois facteurs de redressement prévus aient été inférieurs à 1 (surdénombrement estimé), l'arrondissement n'a pas produit d'estimations gonflées.

Les opérations de redressement ont permis d'ajouter 32,843 noms à la liste des personnes recensées (soit un taux de sous-dénombrement de 8.2%). Si l'on avait utilisé les facteurs de redressement fondés sur l'échantillon, c'est 36,454 noms que l'on aurait ajouté à la liste des personnes recensées (soit un taux de sous-dénombrement de 9.0%). Le lissage a donc réduit de près de 10% le niveau de sous-dénombrement estimé. Cette forte réduction s'explique par le fait que les groupes qui présentaient les estimations de sous-dénombrement les plus élevées avant le lissage étaient ceux qui comptaient le plus grand nombre de personnes.

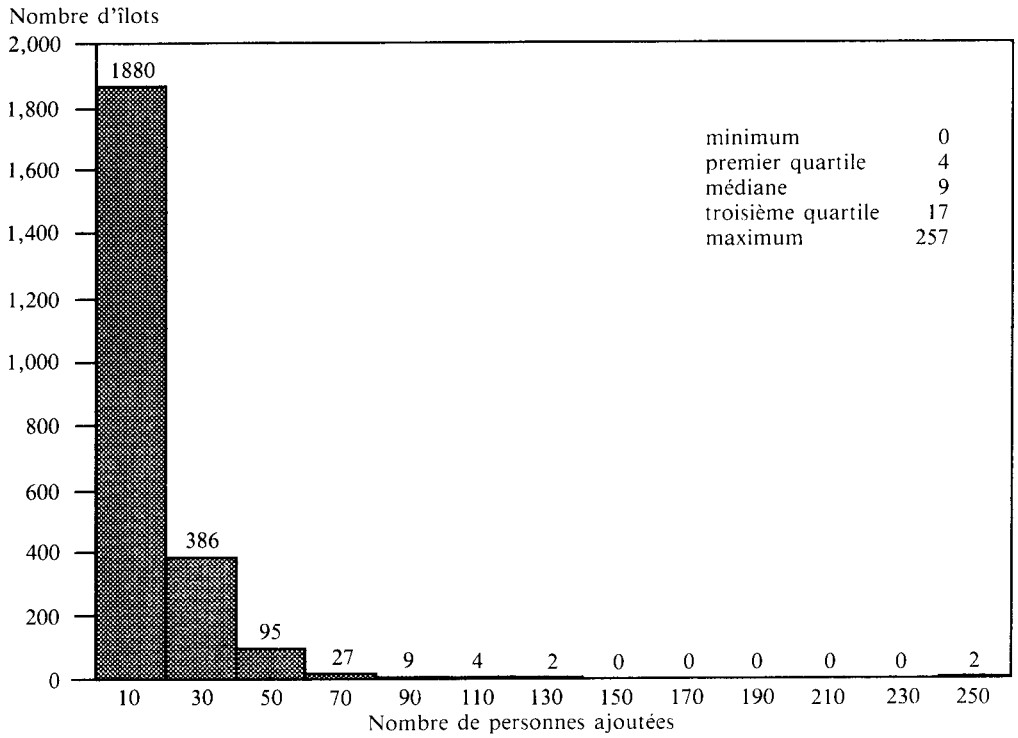

Figure 1. Histogramme et données sommaires concernant le nombre de personnes ajoutées par îlot

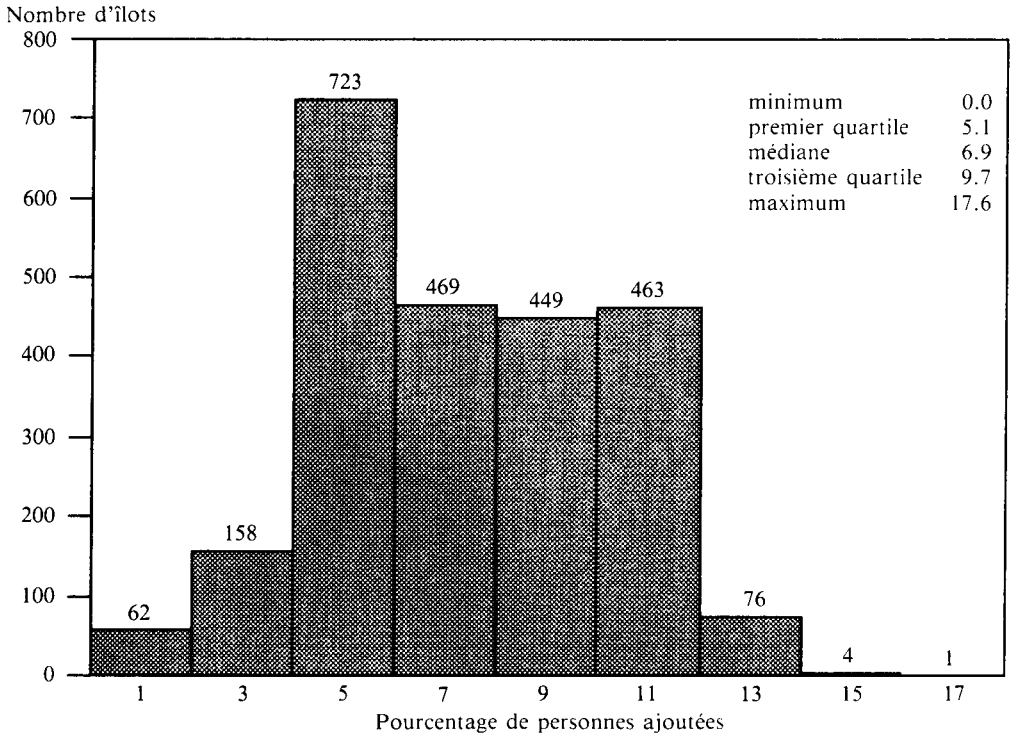

Figure 2. Histogramme et données sommaires concernant le pourcentage de personnes ajoutées par îlot

Pour résumer les opérations de redressement au niveau de l'îlot, nous représentons dans la figure 1 le nombre de personnes qui ont été ajoutées selon le nombre d'îlots concernés et dans la figure 2, le pourcentage de personnes ajoutées selon le nombre d'îlots concernés. Près de 80% des îlots ont reçu moins de 20 personnes. Seulement deux îlots ont reçu plus de 150 personnes. Ces deux îlots étaient assez denses puisqu'ils comptaient chacun environ 2,000 personnes. Plus de 80% des îlots avaient un taux de sous-dénombrement estimé variant de 4% à 12%. Nombre de petits îlots ont reçu un faible pourcentage des personnes additionnelles parce que les estimations avaient été arrondies vers le bas et que cela s'était traduit par une forte variation du pourcentage. Les îlots qui ont reçu le plus fort pourcentage de personnes additionnelles étaient ceux qui comptaient surtout des locataires d'origine hispanique; du reste, ces îlots étaient ceux pour lesquels les facteurs de redressement prévus étaient les plus élevés.

5. CONCLUSION

Cet article nous a permis d'étudier les méthodes, les opérations et les résultats du Test des opérations de redressement (TOR). Ce test visait à vérifier le calendrier et les caractéristiques des opérations de redressement des données du recensement eu égard à la population non dénombrée.

Les résultats du TOR prouvent qu'il est possible de produire des estimations du taux de sous-dénombrement dans des délais raisonnables. En outre, le TOR a été réalisé plus rapidement que n'importe quelle enquête post-censitaire antérieure.

Les résultats du TOR indiquent un taux de sous-dénombrement de 9% pour le recensement d'essai du Central Los Angeles County. Nous avons établi des estimations bivalentes pour 70 catégories formées en fonction de l'origine ethnique et du mode d'occupation, de l'âge et du

sexe. Nous avons lissé ces estimations en leur ajustant un modèle de régression puis nous avons reproduit les estimations ainsi obtenues au niveau de l'îlot. L'utilisation d'un taux de sous-dénombrement estimé pour l'îlot permet l'agrégation de données à n'importe quel niveau au-dessus de l'îlot.

Schenker (1988) et Hogan et Wolter (1988) font une évaluation des opérations et définissent les hypothèses sur lesquelles reposent les estimateurs. Ces deux ouvrages ainsi que le présent article constituent une analyse approfondie des chiffres du recensement et des taux de sous-dénombrement estimés pour le recensement d'essai.

REMERCIEMENTS

L'auteur tient à exprimer sa reconnaissance aux personnes qui ont collaboré à cette étude: Dan Childers, Howard Hogan, Cary Isaki, Matthew Jaro, Jan Jaworski, Charisse Jeffries, Robert O'Brien, Arona Pistiner, Nathaniel Schenker, Maria Urrutia, Debbie Wagner et Kirk Wolter. L'auteur tient aussi à remercier l'arbitre, dont les commentaires ont permis d'améliorer le contenu de cet article.

BIBLIOGRAPHIE

ANOLIK, IRWIN (1988). The 1986 rural post-enumeration survey in East Central Mississippi. Statistical Research Division Report Series. CENSUS/SRD/RR-87/09.

CAUSEY, B.D., COX, L.H., et ERNST, L.R. (1985). Applications of transportation theory to statistical problems. *Journal of the American Statistical Association*, 80, 903-909.

CITRO, CONSTANCE F., et COHEN, MICHAEL L. (1985). *The Bicentennial Census — New Directions for Methodology in 1990*. Washington: National Academy Press.

ERICKSEN, EUGENE P., et KADANE, JOSEPH B. (1985). Estimating the population in a census year — 1980 and beyond. *Journal of the American Statistical Association*, 80, 98-109.

FAY, R.E., PASSEL, J.S., et ROBINSON, J.G. (1988). The coverage of population in the 1980 census. 1980 Census of Population and Housing Evaluation and Research Report PHC80-E4, Washington: U.S. Government Printing Office.

FREEDMAN, D. et NAVIDI, W. (1986). Models for adjusting the census. *Statistical Science*, 1, 3-11.

HOGAN, HOWARD R. (1984). Research plan on adjustment. *Proceedings of the Social Statistics Section American Statistical Association*, 452-457.

HOGAN, HOWARD R., et WOLTER, KIRK M. (1988). Mesure de l'erreur dans une enquête post-censitaire. *Techniques d'enquête*, 14.

ISAKI, CARY T., SCHULTZ, LINDA K., SMITH, PHILIP J., et DIFFENDAL, GREGG J. (1987). Small area estimation research for census undercount-progress report. Dans *Small Area Statistics: An International Symposium*, 219-238. New York: John Wiley and Sons.

JARO, MATTHEW, et CHILDERS, DANNY (1986). Matching the 1985 census of Tampa. Article présenté à la conférence annuelle de l'American Statistical Association. Chicago.

MARKS, ELI S. (1978). The use of dual system estimation in census evaluation. Dans *Developments in Dual System Estimation*, (éd. Karol Krotki), Edmonton: University of Alberta Press.

MULRY, M.H., HOGAN, H.R., WALKER, J.R., CHAPMAN, D.R., EVAUL, J., et MOORE, R.H. (1981). Research proposal for a study of methods for 1990 decennial census coverage evaluation. Rapport technique, U.S., Bureau of the Census, Washington, D.C.

SCHENKER, NATHANIEL (1988). Traitement des données manquantes dans l'estimation de la couverture: le test des opérations de redressement de 1986. *Techniques d'enquête*, 14.

U.S. Bureau of the Census (1979). Evaluation. Dans *POPSTAN: Program Considerations*, Partie A, Chapitre A-13.

WOLFGANG, GLENN (1987). The pre-enumeration survey of the 1986 census of Central Los Angeles County. Article présenté à la conférence annuelle de l'American Statistical Association, San Francisco.

WOLTER, KIRK M. (1986). Some coverage error models for census data. *Journal of the American Statistical Association*, 81, 338-346.

ZASLAVASKY, ALAN M. (1988). Representing local area undercount by reweighting of households. Proceedings of the Bureau of the Census Fourth Annual Research Conference.