

## Méthode alternative pour ajuster les estimations de la Current Population Survey aux chiffres de population

K.R. COPELAND, F.K. PEITZMEIER, et C.E. HOY<sup>1</sup>

### RÉSUMÉ

La Current Population Survey utilise la méthode itérative du quotient dans l'estimation par stratification a posteriori pour redresser les estimations démographiques d'échantillon en fonction d'estimations démographiques fondées sur le recensement. Dans cet article, les auteurs proposent une deuxième méthode, fondée sur les moindres carrés généralisés, et la comparent à la méthode actuelle.

MOTS CLÉS: Moindres carrés généralisés; stratification a posteriori; estimation itérative du quotient.

### 1. INTRODUCTION

La Current Population Survey (CPS) produit des estimations de la population active pour l'ensemble de la population civile hors institutions d'âge actif aux États-Unis à partir d'un échantillon aléatoire mensuel à plusieurs degrés de quelque 60,000 logements. Chaque mois, un échantillon avec renouvellement composé de 8 panels (appelés groupes de renouvellement) de logements est interviewé afin de recueillir des données démographiques et des données sur l'activité pour tous les occupants adultes civils des logements échantillonnés.

Les estimations, sous-agrégées selon des caractéristiques démographiques, sont publiées mensuellement. Des estimations concernant d'autres sous-agrégats de la population (états, familles, anciens combattants, salariés, inactifs, etc.) sont aussi publiées mensuellement, trimestriellement ou annuellement.

On calcule le poids des personnes de l'échantillon en utilisant la probabilité de sélection, un facteur de compensation de la non-réponse et un facteur de correction sous forme de quotient afin de réduire la composante de la variance attribuable au prélèvement d'unités primaires d'échantillonnage. On corrige ensuite les poids ainsi obtenus au moyen d'une méthode d'estimation par stratification a posteriori de manière à rapprocher les estimations démographiques de l'enquête d'estimations démographiques supplémentaires. Enfin, les poids ainsi corrigés sont utilisés dans une méthode d'estimation composite, puis désaisonnalisés afin d'obtenir des estimations nationales (Hanson 1978).

Pour certains sous-domaines de population (familles, salariés, inactifs, revenus de la famille et anciens combattants), le calcul d'estimations détaillées requiert des poids d'échantillon déterminés par des méthodes de correction qui viennent s'ajouter à l'estimation par stratification a posteriori.

La méthode des moindres carrés généralisés (MCG) pourrait peut-être remplacer l'estimation par stratification a posteriori ou servir à intégrer les diverses méthodes de correction de la CPS. On a déjà proposé l'application des MCG dans l'enquête sur les dépenses des consommateurs et des analyses ont été faites à ce sujet (Zieschang 1986).

Nous nous proposons dans cet article de comparer la méthode d'estimation par stratification a posteriori utilisée actuellement dans la CPS (méthode itérative du quotient) et la méthode MCG en nous fondant sur les données de la CPS pour deux mois particuliers (juillet 1983

<sup>1</sup> K.R. Copeland, F.K. Peitzmeier et C.E. Hoy, Division of Statistical Methods, Office of Employment and Unemployment Statistics, Bureau of Labor Statistics, Washington, D.C. 20212, États-Unis.

et juillet 1984). Afin d'apprécier les différences qui pouvaient exister entre les deux méthodes pour cette application, nous nous sommes intéressés aussi bien aux macro-données qu'aux micro-données.

## 2. MÉTHODE D'ESTIMATION PAR STRATIFICATION À POSTERIORI UTILISÉE ACTUELLEMENT DANS LA CPS

Dans la CPS, on utilise actuellement la méthode itérative du quotient (MIQ) pour corriger les poids d'échantillon à l'intérieur d'un groupe de renouvellement de manière à ajuster les estimations démographiques de l'échantillon à des estimations démographiques supplémentaires dans trois catégories distinctes (état, âge/sexe/origine ethnique, âge/sexe/origine raciale).

Deming et Stephan (1940) ont été les premiers à proposer l'application de la MIQ pour la correction des données de tableaux en remplacement de la méthode des moindres carrés. On a montré que la MIQ produisait de meilleurs estimateurs asymptotiquement normaux (MEAN) dans un échantillonnage aléatoire simple et qu'elle minimisait, comme nous le verrons dans la sous-section 4.2, les corrections apportées aux poids d'échantillon (Ireland et Kullback 1968). De plus, même si elle produit des estimations biaisées, la MIQ peut parfois réduire l'erreur quadratique moyenne des estimations d'enquête. C'est le rôle que l'on attribue à la MIQ dans la CPS (Hanson 1978).

En ce qui a trait à la CPS, la MIQ vise à corriger les chiffres d'échantillon  $\{n_{ijk}\}$  découlant des étapes de pondération antérieures afin de produire des chiffres d'échantillon rajustés  $\{\tilde{n}_{ijk}\}$  à la condition que

$$(A) \quad \sum_{j,k} \tilde{n}_{ijk} = m_{i..}$$

$$(B) \quad \sum_{i,k} \tilde{n}_{ijk} = m_{.j.}$$

$$(C) \quad \sum_{i,j} \tilde{n}_{ijk} = m_{..k}$$

soient satisfaites simultanément,

où  $i$  = état ( $i = 1, \dots, 51$ ),

$j$  = âge/sexe/origine ethnique ( $j = 1, \dots, 16$ ),

$k$  = âge/sexe/origine raciale ( $k = 1, \dots, 70$ ),

$m_{i..}$  = estimation supplémentaire relative à l'état,

$m_{.j.}$  = estimation supplémentaire relative à la combinaison âge/sexe/origine ethnique,

$m_{..k}$  = estimation supplémentaire relative à la combinaison âge/sexe/origine raciale.

La MIQ permet de corriger proportionnellement, en se servant de quotients, les données d'échantillon selon les trois catégories de critères (c'est-à-dire, état, âge/sexe/origine ethnique et âge/sexe/origine raciale) par étapes successives.

(1) Correction par quotient selon l'état:

$$n_{ijk}^{(1,1)} = (m_{i..} / n_{i..}) n_{ijk} = a_i^{(1)} n_{ijk}.$$

2) Correction par quotient selon la combinaison âge/sexe/origine ethnique:

$$\begin{aligned} n_{ijk}^{(1,2)} &= (m_{.j} / n_{.j}^{(1,1)}) n_{ijk}^{(1,1)} = b_j^{(1)} n_{ijk}^{(1,1)} \\ &= a_i^{(1)} b_j^{(1)} n_{ijk}. \end{aligned}$$

3) Correction par quotient selon la combinaison âge/sexe/origine raciale:

$$\begin{aligned} n_{ijk}^{(1,3)} &= (m_{..k} / n_{..k}^{(1,2)}) n_{ijk}^{(1,2)} = d_k^{(1)} n_{ijk}^{(1,2)} \\ &= a_i^{(1)} b_j^{(1)} d_k^{(1)} n_{ijk}, \end{aligned}$$

où  $n_{i..}$  = total de ligne pour l'échantillon

$n_{.j.}$  = total de colonne pour l'échantillon

$n_{..k}$  = total de couche pour l'échantillon.

L'exécution des trois étapes ci-dessus correspond à une itération de la méthode itérative du quotient. On reprend les trois étapes en substituant à chaque fois la valeur de  $n_{ijk}^{(h,3)}$  (donnée-échantillon corrigée issue de la troisième étape de la  $h$ -ième itération) à  $n_{ijk}$  dans l'étape 1) jusqu'à ce qu'on ait effectué 6 itérations. (Le nombre d'itérations effectuées dans la CPS a été établi en fonction des propriétés de convergence de la MIQ pour cette enquête et des gains relatifs réalisés selon le nombre d'itérations.) La valeur finale  $\{n_{ijk}^{(6,3)}\}$  est définie comme  $\{\tilde{n}_{ijk}\}$ .

Le facteur de correction des poids d'échantillon pour les enregistrements de la cellule  $\{ijk\}$  est

$$\begin{aligned} F_{ijk} &= n_{ijk}^{(6,3)} / n_{ijk} \\ &= \prod_{h=1}^6 a_i^{(h)} b_j^{(h)} d_k^{(h)}. \end{aligned}$$

Pour obtenir les poids corrigés, on multiplie les poids d'échantillon observés avant l'application de la MIQ par le facteur  $F_{ijk}$  approprié.

### 3. APPLICATION DES MCG DANS LA CPS

La méthode des moindres carrés généralisés (MCG) permet de corriger les poids d'échantillon découlant des étapes de pondération antérieures en minimisant les carrés pondérés des corrections, sous réserve que les poids corrigés satisfassent à une série de contraintes "de contrôle" linéaires. C'est exactement ce à quoi Deming et Stephan cherchaient à s'attaquer lorsqu'ils ont mis au point la MIQ. Comme celle-ci, la méthode MCG produit des MEAN dans certaines conditions, en l'occurrence lorsque toutes les cellules sont non nulles (Neyman 1949). Par définition, la méthode MCG minimise les corrections apportées aux poids d'échantillon par l'intermédiaire d'une mesure de proximité (voir sous-section 4.2).

En ce qui a trait à la CPS, chaque dimension qui définit une série de chiffres de population dans la stratification a posteriori courante définira aussi une série de contraintes linéaires pour la méthode MCG. La fonction à minimiser est

$$\begin{aligned} f(\underline{F}) &= (\underline{F} - \underline{P})' P_0^{-1} (\underline{F} - \underline{P}) \\ &= \sum_i (W_{2i} - W_{1i})^2 / W_{1i}, \end{aligned}$$

à la condition que  $X' \underline{F} = \underline{N}$ ,

où  $\underline{F}$  = vecteur ( $n \times 1$ ) des poids finals calculés ( $W_{2i}$ ) pour chacune des  $n$  personnes de l'échantillon,

$\underline{P}$  = vecteur ( $n \times 1$ ) des poids des personnes de l'échantillon avant la stratification a posteriori ( $W_{1i}$ ),

$\underline{P}_0$  = matrice diagonale ( $n \times n$ ) dont la diagonale est formée des  $W_{1i}$ ,

$\underline{X}$  = matrice descriptive ( $n \times k$ ), où les lignes correspondent aux personnes de l'échantillon et les colonnes aux cellules de contrôle. La matrice est formée de zéros (0) et de uns (1), qui indiquent les catégories de contrôle appropriées pour chacune des  $n$  personnes de l'échantillon,

$\underline{N}$  = vecteur ( $k \times 1$ ) des estimations démographiques supplémentaires, qui correspondent aux colonnes de  $\underline{X}$ . Ces estimations servent aussi à la MIQ dans la CPS.

Les colonnes de  $\underline{X}$  doivent être linéairement indépendantes pour qu'il existe un inverse de la matrice ( $\underline{X}' \underline{P}_0 \underline{X}$ ). Lors de la formation des matrice  $\underline{X}$  et  $\underline{N}$  pour la CPS, les 137 cellules de contrôle utilisées pour la MIQ (état, âge/sexe/origine ethnique, âge/sexe/origine raciale) ont été réduites à  $k = 132$  cellules qui étaient linéairement indépendantes.

La seule solution de  $\underline{X}' \underline{F} = \underline{N}$  qui minimise  $f(\underline{F})$  est (voir Luery 1986)

$$\underline{F} = \underline{P} + \underline{P}_0 \underline{X} (\underline{X}' \underline{P}_0 \underline{X})^{-1} (\underline{N} - \underline{X}' \underline{P})$$

Bien qu'il ne soit pas nécessaire que les éléments de  $\underline{F}$  soient positifs, ils l'étaient tous dans ce cas sans que l'on ait eu à poser des conditions additionnelles. Huang et Fuller (1978) et Zieschang (1986), entre autres, analysent des méthodes qui permettent d'obtenir des poids non négatifs dans ce contexte.

## 4. RÉSULTATS

### 4.1 Macro-données

#### a. Estimations

Les estimations de la population active ont été établies pour plusieurs groupes démographiques pour juillet 1983 et juillet 1984 au moyen des poids finals calculés par la MIQ et les MCG. On a aussi calculé les erreurs types pour les deux méthodes au moyen d'un estimateur de la méthode des groupes aléatoires défini comme suit (Wolter 1985):

$$\sum_{k=1}^8 (8Y_k - \hat{Y})^2 / 56,$$

où  $Y_k$  = somme des poids des enregistrements de l'échantillon pour le  $k$ -ième groupe de renouvellement avec la caractéristique  $Y$ ,

$\hat{Y}$  = somme des  $Y_k$ .

On a utilisé cet estimateur de variance, même s'il ne tient pas compte du plan de sondage à plusieurs degrés de la CPS, parce que le fichier de micro-données à grande diffusion de la CPS ne contenait aucune information sur le plan de sondage.

On a calculé des différences relatives pour les estimations de niveau et les estimations d'erreur type. La formule de la différence relative était

$$(Y_{GLS} - Y_{RRE}) / Y_{RRE},$$

où  $Y_{MIQ}$  = estimation de  $Y$  fondée sur les poids calculés selon la MIQ,

$Y_{MCG}$  = estimation de  $Y$  fondée sur les poids calculés selon la méthode des MCG.

**Tableau 1**  
Estimations de la population active selon la combinaison sexe/origine raciale ou ethnique

		1983				1984			
		MCG		(MCG-MIQ)/ MIQ		MCG		(MCG-MIQ)/ MIQ	
		Total (milliers)	S.E. (milliers)	Total (%)	S.E. (%)	Total (milliers)	S.E. (milliers)	Total (%)	S.E. (%)
<b>Total</b>									
Total	Occupées	103516	403	0.00	-0.14	107535	352	-0.01	1.12
	En chômage	10669	221	-0.04	-0.75	8765	118	-0.06	-0.21
	<i>Taux</i>	<i>9.34%</i>	<i>0.19%</i>	<i>-0.04</i>	<i>-0.56</i>	<i>7.54%</i>	<i>0.09%</i>	<i>-0.05</i>	<i>0.27</i>
	Inactives	59938	373	0.01	-0.68	60080	419	0.02	0.41
Personnes de race blanche	Occupées	91338	344	0.00	-0.33	94417	274	0.00	0.70
	En chômage	7928	236	0.00	-0.27	6282	120	0.00	-0.14
	<i>Taux</i>	<i>7.99%</i>	<i>0.23%</i>	<i>0.00</i>	<i>-0.26</i>	<i>6.24%</i>	<i>0.10%</i>	<i>0.00</i>	<i>-0.16</i>
	Inactives	51915	340	0.00	-0.36	51700	358	0.00	0.39
Personnes de race noire	Occupées	9871	69	0.06	-3.44	10371	98	0.02	0.17
	En chômage	2434	68	-0.12	-1.07	2202	60	-0.03	1.41
	<i>Taux</i>	<i>19.78%</i>	<i>0.55%</i>	<i>-0.14</i>	<i>-1.60</i>	<i>17.51%</i>	<i>0.42%</i>	<i>-0.04</i>	<i>1.49</i>
	Inactives	6628	26	-0.04	-1.47	6765	109	-0.02	0.09
Personnes d'origine hispanique	Occupées	6132	73	-0.03	-0.59	6607	102	-0.03	1.90
	En chômage	920	79	-0.05	-0.29	786	70	-0.08	-0.03
	<i>Taux</i>	<i>13.04%</i>	<i>1.10%</i>	<i>-0.02</i>	<i>-0.33</i>	<i>10.63%</i>	<i>0.96%</i>	<i>-0.05</i>	<i>0.35</i>
	Inactives	3760	31	0.05	-0.39	3786	73	0.04	1.02
<b>Hommes</b>									
Total	Occupées	58985	147	0.00	-1.58	61045	188	0.00	1.74
	En chômage	5980	134	-0.05	-0.88	4682	79	-0.02	0.77
	<i>Taux</i>	<i>9.20%</i>	<i>0.19%</i>	<i>-0.05</i>	<i>-0.79</i>	<i>7.12%</i>	<i>0.11%</i>	<i>-0.02</i>	<i>1.30</i>
	Inactives	17495	178	0.01	-1.81	17840	214	0.02	0.64
De race blanche	Occupées	52674	482	0.00	0.42	54261	111	0.00	0.34
	En chômage	4484	131	0.01	-0.49	3394	93	0.01	-0.12
	<i>Taux</i>	<i>7.84%</i>	<i>0.21%</i>	<i>0.00</i>	<i>-0.47</i>	<i>5.89%</i>	<i>0.15%</i>	<i>0.01</i>	<i>-0.13</i>
	Inactives	14985	160	-0.02	-0.40	15077	150	0.00	0.16
De race noire	Occupées	5047	56	0.07	-1.70	5263	84	0.01	-0.50
	En chômage	1300	45	-0.20	-1.87	1137	33	0.08	1.12
	<i>Taux</i>	<i>20.49%</i>	<i>0.71%</i>	<i>-0.21</i>	<i>-2.02</i>	<i>17.76%</i>	<i>0.51%</i>	<i>0.05</i>	<i>0.94</i>
	Inactives	2097	40	-0.04	-0.13	2236	88	-0.07	-0.48
D'origine hispanique	Occupées	3781	48	0.01	-0.86	4064	79	-0.02	1.29
	En chômage	534	45	-0.16	-0.83	451	41	-0.05	0.51
	<i>Taux</i>	<i>12.38%</i>	<i>0.99%</i>	<i>-0.15</i>	<i>-0.89</i>	<i>9.99%</i>	<i>0.95%</i>	<i>-0.03</i>	<i>0.66</i>
	Inactives	981	42	0.00	-0.42	964	57	0.07	1.40
<b>Femmes</b>									
Total	Occupées	44531	320	-0.01	-0.01	46490	194	-0.01	1.48
	En chômage	4689	107	-0.04	-0.19	4083	88	-0.10	-1.22
	<i>Taux</i>	<i>9.53%</i>	<i>0.23%</i>	<i>-0.03</i>	<i>-0.02</i>	<i>8.07%</i>	<i>0.16%</i>	<i>-0.09</i>	<i>-0.80</i>
	Inactives	42443	287	0.01	-0.26	42240	217	0.02	0.34
De race blanche	Occupées	38664	315	0.00	-0.29	40156	191	0.00	0.66
	En chômage	3444	115	-0.01	0.16	2888	68	0.00	-0.32
	<i>Taux</i>	<i>8.18%</i>	<i>0.28%</i>	<i>-0.01</i>	<i>0.11</i>	<i>6.71%</i>	<i>0.15%</i>	<i>0.00</i>	<i>-0.34</i>
	Inactives	36929	283	0.01	-0.32	36623	214	0.00	0.53
De race noire	Occupées	4824	57	0.05	0.56	5108	50	0.02	1.69
	En chômage	1134	46	-0.02	0.07	1065	46	-0.14	-0.62
	<i>Taux</i>	<i>19.03%</i>	<i>0.80%</i>	<i>-0.06</i>	<i>0.08</i>	<i>17.25%</i>	<i>0.67%</i>	<i>-0.13</i>	<i>-0.63</i>
	Inactives	4531	24	-0.04	2.99	4529	59	0.01	1.49
D'origine hispanique	Occupées	2350	44	-0.08	-0.46	2543	38	-0.05	3.04
	En chômage	385	41	0.10	0.51	335	34	-0.13	-0.62
	<i>Taux</i>	<i>14.08%</i>	<i>1.46%</i>	<i>0.16</i>	<i>0.57</i>	<i>11.64%</i>	<i>1.18%</i>	<i>-0.07</i>	<i>-0.11</i>
	Inactives	2778	33	0.07	-0.87	2822	27	0.03	0.13

Comme l'indiquent les données du tableau 1, aucune différence ou tendance notable ne se dégage des estimations pondérées de la population active ou des estimations de l'erreur type calculées par l'une et l'autre des deux méthodes lorsque ces estimations sont sous-agrégées en fonction du sexe et de l'origine raciale ou ethnique.

En ce qui concerne les estimations de la population active selon la combinaison sexe/origine raciale ou ethnique, les différences relatives estimées (en valeur absolue) étaient toutes inférieures à 0.3% (ce qui est bien au-dessous des CV estimés de chaque estimation). Pour la majorité de ces estimations, notamment celles ayant trait à l'ensemble des personnes de sexe masculin ou de sexe féminin ou aux personnes de race blanche, la différence relative en valeur absolue était inférieure à 0.1%.

Pour un bon nombre de caractéristiques, le signe de la différence relative a changé de 1983 à 1984; l'écart entre les estimations produites par les deux méthodes ne semble donc pas suivre une tendance particulière.

En ce qui concerne les estimations de l'erreur type pour les estimations nationales de la population active, les différences relatives en valeur absolue étaient toutes inférieures à 1.9% pour la population totale, à 0.7% pour les personnes de race blanche, à 3.5% pour les personnes de race noire et à 3.1% pour les personnes d'origine hispanique.

#### **b. Indices de l'effet du nombre de mois d'inclusion dans l'échantillon**

De nombreuses études montrent que les estimations calculées à l'aide des poids finals de la CPS sont entachées d'un biais relatif qui varie selon le nombre de mois d'inclusion du groupe de renouvellement dans l'échantillon (Bailar 1975). Des indices de l'effet du nombre de mois d'inclusion dans l'échantillon

$$I_k = (8Y_k / \hat{Y}) \times 100,$$

ont été calculés pour juillet 1983 et juillet 1984 d'après les estimations calculées par la MIQ et les MCG.

Dans les deux cas, on a obtenu des indices virtuellement identiques pour la population active selon l'origine raciale, selon le sexe et selon l'origine ethnique.

### **4.2 Micro-données**

#### **a. Correction des poids d'échantillon**

La MIQ aussi bien que les MCG minimisent une mesure de proximité des poids d'échantillon (poids initial vs poids corrigé). Dans le cas de la MIQ, la mesure est (Ireland et Kullback 1968)

$$M_A = \sum_i W_{2i} \ln(W_{2i} / W_{1i}).$$

Pour les MCG, la mesure est (Luery 1986)

$$M_B = \sum_i (W_{2i} - W_{1i})^2 / W_{1i}$$

où  $W_{1i}$  = poids de l'enregistrement  $i$  de l'échantillon avant correction,

$W_{2i}$  = poids de l'enregistrement  $i$  de l'échantillon après correction.

La comparaison des mesures de proximité (voir le tableau 2) donne des résultats intéressants et parfois étonnants. Ainsi, la MIQ a produit des valeurs moindres pour les deux mesures. La méthode des MCG avait tendance à produire des valeurs moindres pour certains sous-groupes, notamment les personnes de race noire et les personnes d'origine hispanique. Il convient de souligner que l'écart entre les valeurs produites par les deux méthodes était presque toujours inférieur à 1%.

**Tableau 2**  
 Comparaison des mesures de proximité fondée  
 sur 8 groupes de renouvellement pour chaque année  
 (nombre de GR pour lesquels MIQ < MCG)

	$M_A$		$M_B$	
	1983	1984	1983	1984
Total	8	8	4	7
De race blanche	7	7	3	4
De race noire	3	3	1	1
D'origine hispanique	0	0	0	0
Hommes	2	7	1	5
Femmes	8	8	8	8

Bien que  $M_B$  devrait normalement être minimisée par la méthode des MCG, la valeur de  $M_B$  fondée sur les poids calculés selon cette méthode pour l'échantillon total était plus élevée que la valeur de  $M_B$  fondée sur les poids calculés selon la MIQ dans 11 cas sur 16.

En cherchant à expliquer cette contradiction apparente, nous avons remarqué qu'après six itérations, les estimations produites par la MIQ ne concordaient toujours pas avec les chiffres de population pour l'âge, le sexe et l'origine ethnique. Le degré de "non-convergence" est toutefois *très faible* (moins de 1.0% pour toutes les catégories de contrôle). Cependant, compte tenu de l'écart entre les valeurs de  $M_B$  fondées respectivement sur la MIQ et les MCG, il suffirait de modifier les poids calculés selon la MIQ dans une proportion de 0.1 à 0.2% pour obtenir le résultat inverse. En soumettant la MIQ à 15 itérations, qui n'ont pas suffi néanmoins à assurer la convergence, nous avons pu constater que les résultats observés pour  $M_B$  pouvaient être attribuables à la légère non-convergence de la MIQ. (Il convient de souligner que la méthode des MCG minimise  $M_B$  pour la série de méthodes de correction qui produisent des estimations conformes aux chiffres de population. Comme les estimations produites par la MIQ lors de la CPS ne concordent pas avec les chiffres de population, elle ne fait pas partie de la catégorie précitée.)

Bien qu'une méthode de correction comme la MIQ ou les MCG puisse minimiser une mesure de proximité pour l'échantillon global, il n'est pas dit qu'elle fera la même chose pour des sous-agrégats de l'échantillon qui ont été comparés à des chiffres de population (par exemple, personnes de race noire, personnes d'origine hispanique, personnes de sexe masculin). Compte tenu de l'utilisation de chiffres de population et du fait que la mesure globale de proximité est minimisée, il semblerait souhaitable d'avoir une méthode de correction qui produise aussi de faibles mesures de proximité pour les sous-agrégats. La méthode des MCG a produit de telles mesures dans presque tous les groupes de renouvellement pour les personnes d'origine hispanique, dans de nombreux groupes de renouvellement pour les personnes de race noire et dans plusieurs groupes de renouvellement pour les personnes de race blanche et de sexe masculin.

#### b. Comparaison des corrections

La MIQ et la méthode MCG définissent toutes deux des facteurs de correction pour les cellules formées par l'intersection des contraintes marginales. Le même facteur s'applique à tous les enregistrements d'une cellule. Afin de comparer les corrections apportées aux poids par l'une ou l'autre des deux méthodes, nous avons comparé les facteurs définis par chacune des méthodes pour chaque enregistrement de l'échantillon à l'aide du rapport suivant:

$$RRE/GLS = [(W_{2i}/W_{1i})_{RRE}] / [(W_{2i}/W_{1i})_{GLS}] .$$

**Tableau 3**  
 Comparaison des corrections apportées par la MIQ et les MCG, 1984

Catégorie de contrôle	Intervalle de taux de couverture	Proportion de l'échantillon total	Proportion de l'échantillon pour laquelle MIQ/MCQ < 0.95 ou > 1.05	Proportion des enregistrements dans l'intervalle pour lesquels MIQ/MCQ < 0.95 ou > 1.05
Âge/sexe/ origine raciale	< 0.7	0.007	0.057	0.219
	0.7-0.8	0.022	0.116	0.136
	0.8-0.9	0.241	0.147	0.019
	0.9-1.1	0.699	0.504	0.019
	1.1-1.2	0.021	0.069	0.084
	> 1.2	0.010	0.106	0.275
Âge/sexe/ origine ethnique	< 0.7	0.010	0.078	0.198
	0.7-0.8	0.014	0.032	0.058
	0.8-0.9	0.106	0.135	0.033
	0.9-1.1	0.869	0.741	0.022
	1.1-1.2	0.001	0.007	0.202
	> 1.2	0.001	0.007	0.373
État	< 0.7	0.056	0.068	0.031
	0.7-0.8	0.111	0.180	0.042
	0.8-0.9	0.278	0.325	0.030
	0.9-1.1	0.479	0.342	0.018
	1.1-1.2	0.026	0.009	0.009
	< 1.2	0.049	0.077	0.040

Ce rapport décrit la relation entre les corrections apportées au poids d'une personne de l'échantillon par la MIQ d'une part et par la méthode MCG d'autre part. Pour les besoins de la comparaison, nous dirons qu'un rapport inférieur à 0.95 ou supérieur à 1.05 représente un écart significatif entre les corrections apportées par l'une et l'autre des deux méthodes.

Pour chaque série de chiffres de population, nous avons calculé des ratios  $E/C$  (taux de couverture), où  $E$  est l'estimation d'échantillon fondée sur les poids des personnes de l'échantillon avant la stratification a posteriori et  $C$  est le chiffre de population.

Pour chaque catégorie de contrôle (état, âge/sexe/origine ethnique, âge/sexe/origine raciale), nous avons classé les enregistrements de l'échantillon selon le taux de couverture. Le tableau 3 donne la répartition de l'échantillon selon l'intervalle de taux de couverture et la valeur du rapport  $MIQ/MCG$ , de même que la proportion des enregistrements dans chaque intervalle de taux de couverture pour lesquels le rapport  $MIQ/MCG$  est inférieur à 0.95 ou supérieur à 1.05.

Les données de ce tableau indiquent que pour chaque catégorie de contrôle, la MIQ et la méthode MCG étaient plus susceptibles de produire des corrections divergentes dans le cas d'enregistrements tirés de groupes de population qui avaient été surdénombrés ou sousdénombrés jusqu'à un certain point dans l'enquête (c'est-à-dire pour lesquels le taux de couverture n'était pas près de 1) que dans le cas d'enregistrements tirés de groupes de population qui avaient été dénombrés normalement.

### 4.3 Ressources informatiques

L'application des deux méthodes a été exécutée sur un Système 370 d'IBM au National Institute of Health au moyen de PROC MATRIX du système SAS. La préparation des fichiers



et la pondération ont exigé un temps d'utilisation du processeur environ trois fois plus long dans le cas de la méthode MCG que dans le cas de la MIQ. La méthode MCG a nécessité aussi plus de temps pour la mise en mémoire des fichiers. (Les matrices utilisées pour la CPS ont des dimensions très appréciables, par exemple  $\underline{P}$ ,  $P_0$ ,  $X$ , et  $\underline{N}$  comptent environ 14,000 lignes pour chaque groupe de renouvellement.)

## 5. RÉSUMÉ ET CONCLUSIONS

Cet article avait pour but de comparer la MIQ et la méthode MCG dans le cadre de la CPS tant au point de vue des macro-données qu'au point de vue des micro-données.

En ce qui concerne les macro-données, les deux méthodes produisent des estimations comparables.

Les mesures de proximité ont permis de constater que les corrections apportées aux poids d'échantillon pour respecter les contraintes de contrôle étaient légèrement inférieures dans le cas de la MIQ. En revanche, celle-ci avait tendance à produire des mesures de proximité légèrement plus élevées (par rapport aux MCG) pour les sous-agrégats de minorités ethniques. Là où les deux méthodes diffèrent le plus l'une de l'autre a trait aux corrections qui sont apportées aux groupes de population qui sont surdénombrés ou sousdénombrés.

Cette analyse nous permet de déduire que la MIQ exige moins de temps machine que la méthode MCG pour la correction du second degré de la CPS.

## REMERCIEMENTS

Les auteurs tiennent à exprimer leur reconnaissance à Fritz Scheuren, qui a révisé la version originale de cet article, de même qu'aux arbitres et au rédacteur associé qui, par leurs commentaires très utiles, ont contribué à améliorer la qualité de cet article.

## BIBLIOGRAPHIE

- BAILAR, B. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70, 23-30.
- DEMING, W.E., et STEPHAN, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.
- HANSON, R.H. (1978). The Current Population Survey design and methodology. Technical Paper 40, U.S. Bureau of the Census.
- HUANG, E.T., et FULLER, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Section on Social Statistics, American Statistical Association*, 300-305.
- IRELAND, C.T., et KULLBACK, S. (1968). Contingency tables with given marginals. *Biometrika*, 55, 179-188.
- LUERY, D. (1986). Weighting sample survey data under linear constraints on the weights. *Proceedings of the Section on Social Statistics, American Statistical Association*, 325-350.
- NEYMAN, J. (1949). Contribution to the Theory of the  $X^2$  Test. Dans *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, (éd. J. Neyman), Berkeley: University of California Press, 239-273.
- WOLTER, K.M. (1985). *Introduction to Variance Estimation*. New York: Springer-Verlag.
- ZIESCHANG, K.D. (1986). A generalized least squares weighting system for the Consumer Expenditure Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 64-71.